Designing for Moral Identity in Information Technology

Designing for Moral Identity in Information Technology

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.ir. K.C.A.M. Luyben
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op vrijdag 25 juni 2010 om 10.00 uur

door Noëmi Laura Johanna Louise HUITS
doctorandus in de filosofie
doctorandus in de filosofie van de informatietechnologie

geboren te Eindhoven (Nederland)

Dit proefschrift is goedgekeurd door de promotoren:
Prof. dr. M.J. van den Hoven
Prof. dr. Y.H. Tan

Copromotor:
Dr. S. Roeser

Samenstelling promotiecommissie
Rector Magnificus, voorzitter
Prof. dr. M.J. van den Hoven, Technische Universiteit Delft, promotor
Prof. dr. Y.H. Tan, Technische Universiteit Delft, promotor
Dr. S. Roeser, Technische Universiteit Delft, copromotor

Prof. dr. J. Perry, Stanford University
Prof. dr. E.J. Koops, Universiteit van Tilburg
Prof. dr. R.L. Lagendijk, Technische Universiteit Delft
Dr. D.W. Shoemaker, Tulane University

Huits, Noëmi Laura Johanna Louise
Designing for Moral Identity in Information Technology

ISBN: 978-90-386-2269-9
ISSN: 1574-941

Contents

Acknowledgements

For you, the reader, these may be the first words you read of this thesis. For me they are the last. In a few weeks, 'having completed a PhD' will be one of my identity tags. It will be one of the ways I can be characterized. I now wonder, in the mode of this thesis, what will be the impact of this characterization? How might this affect me, in a practical, but also in a moral sense?...

On a more personal level, completing a PhD was yet another one of my ambitions, of challenging myself. And a challenge it was: these past years have been a test on my perseverance, creativity, precision, to name only a few. It was enjoyable and strenuous at the same time: Especially in the final phase, I felt torn between the need and the desire to finish, and the opposing desire to get to know more and to enjoy the luxury of learning and reflecting about this topic for much longer.

In any case I now look back at an exciting time of my life. There have been many who have in some way contributed to the realization and success of this PhD, and I am grateful. Here I'd like to acknowledge the most important ones.

First and foremost I am heavily indebted to my supervisors Jeroen van den Hoven, Yao-Hua Tan and Sabine Roeser. Jeroen van den Hoven has never failed to challenge me both intellectually and personally. He has encouraged me to engage in this field and stimulated me to greater efforts. Jeroen, thank you for coaching me all these years. My late supervisor, Rene Wagenaar, also deserves special thanks for his involvement and dedication in the first years of my PhD. His sudden death was a great loss in many ways. Thanks to Yao-Hua Tan for taking up his role. Finally, I thank Sabine for her great reliability and stimulating enthusiasm, and for our sociable, amusing talks.

I also thank all of my colleagues for their support and for making the workplace in Delft a pleasant place to be. I'd especially like to mention David Koepsell, Richard Heersmink, Christian Dettweiler and Ibo van de Poel for proofreading some parts of my work. Moreover, I thank my 'external colleagues' Paul Sollie and Michael Zimmer: I've enjoyed collaborating with you very much.

On numerous occasions during my PhD I've had the privilege of going abroad, both for conferences and visits of another kind. Especially my visits to CAPPE (ANU) in Australia, NYU and UVA in the United States and the NSF-sponsored summer school on Values in Design in Santa Clara are treasured. For

these opportunities I'd explicitly like to thank TU Delft, Jeroen van den Hoven and Peter Kroes, Seumas Miller and John Weckert (ANU), Helen Nissenbaum (NYU), Deborah Johnson (UVA), Geoffrey Bowker and the Dutch Research Council. In addition to these visits another memorable experience was the three-day international conference I organized on Ethics, Technology, and Identity. For making this possible I am grateful to TU Delft, the 3TU.Centre for Ethics and Technology, to the Royal Dutch Academy of Sciences; but of course also to the participants and to my keynote speakers Oscar Gandy, David Velleman, Robin Dillon and David Shoemaker.

To my paranymphs Thomas and Nadine: Many thanks for being there for me on this special occasion. Thomas, you have always inspired me by being inquisitive and enterprising. And Nadine, you have always supported me in life choices and struggles, regardless of the outcome of these episodes. Besides that you are always fun to be with, with your enduring enthusiasm and curiosities.

To be sure my family deserves special thanks. Robin, Yannick, Maxime, and Quésar, I cannot imagine living in greater wealth. I am also grateful to my parents, Loes and Martin, and to my parents-in-law, Nelleke and Hans, for their continuous interest and support. And Juul, outlining this part of our lives here is impossible. It was intense. I can only express my deepest gratitude and leave the last words for you.

NMH

1 Introduction

Suppose there is a world in which human beings are primarily conceived through genetic engineering. As a result, DNA plays a crucial role in determining social class and the associated opportunities. In this world lives Vincent, a young man, exceptional because he was conceived without the aid of technology. By means of screening it is determined that he has a life expectancy of 30.2 years, and the potential of developing some highly uncommon, nearly eradicated, physical dysfunctions. Vincent dreams of becoming an astronaut, which is unfortunately something that is only reserved for human beings with a top-class, 'A'-rated genetic profile.

But then he meets Jerome, a former swimming star who has just such an exceptional genetic profile. Jerome however, is paralyzed after a suicide attempt, which he committed because he could not meet the expectations associated with his profile: He 'only' won a silver medal in an important competition. His disability is unknown, as he hides in his home since the incident. Therefore Vincent can assume his identity with no one the wiser. He uses Jerome's genetic profile to enter the institute for education to become an astronaut. His interview consists entirely of a urine test and he is directly accepted. At the institute, Vincent is subjected to daily DNA sampling, for example in order to enter the building. For this he uses Jerome's blood, tissue, and urine samples. In addition, he has to make sure he rids himself of his own skin flakes and possible loose body hairs. This way he can pursue his career and quickly becomes one of the company's ace celestial navigators.

This story, drawn from the movie *Gattaca*,[1] centers on the irony of someone with a 'perfect' profile, Jerome, failing to succeed despite being given every opportunity, and the one with the imperfect profile, Vincent, who overcomes his (potential) deficiencies and imposed constraints through spirit and force of will. It demonstrates the opposition between one's 'moral identity' and the profile - or identity - he or she has acquired. Although it is fiction, it aptly demonstrates the area of tension that is the main focus of this research. What is more, its setting contains elements that are likewise apparent in modern society, such as the use

[1] *Gattaca* (1997) is a movie written and directed by Andrew Niccol.

of information technology for screening, selection on the basis of profiles,[2] and the associated probabilistic organization of society, i.e. a society in which probabilities and risk calculations are used as input for decisions and policy.

A recent example of profile-based selection, or the 'targeting' of customers, concerns the use of algorithms to select potential voters by means of so-called 'Voter micro-targeting.' This was used during the U.S. presidential elections of 2004 (see Lundry: 2004).[3] Voter micro-targeting entails the selection not only of potential voters who are likely to vote, but also their likelihood to support a particular party dependent on certain issues that are salient to them. With the help of data mining techniques, people are selected and approached only because they are believed by campaign teams to be able to be persuaded to vote for a candidate by focusing on one single issue, leaving competing concerns aside.

Another topical illustration concerns the attuning of (shop or service) supplies to (potential) customers. Consider pre-selections by Amazon.com, cable-TV providers, or insurance companies. In some cases, what is preselected is only the presentation of supplies; in other cases competing supplies or services aren't even an option. What these cases have in common is that the range of opportunities for each subject is confined as a consequence of the use of (identity management and) profiling techniques. This leads to the well-known advantages such as scale, efficiency and service level. But as I argue in this thesis, using information technology for identification, selection, and classification also affects the way identity is perceived in a morally relevant sense.

The mediation of our daily lives by information technology (both online and offline) cannot have escaped anyone's notice. The institutions we deal with on a daily basis are increasingly automated or taking place online; public administration agencies, such as tax authorities, or financial institutions such as banks are cases in point. With the advent of information technologies, administrative duties have become easier to manage. The ease of storing, retrieving, processing, and exchanging information have doubtlessly accelerated, facilitated, and improved both private and public services. Access to these

---

[2] *Gattaca* primarily concerns the use of genetic (DNA) material for profiling and selection, whereas current profiling is rather based on behavior.

[3] Lundry is Research Director for TargetPoint Consulting, the firm that helped steer the Bush re-election campaign. On voter discrimination, see also Philip N. Howard, *New Media Campaigns and the Managed Citizen*, New York: Cambridge University Press, 2005.

services is (primarily) personal, and regulated by means of establishing one's identity, for example by means of 'user identities' and passwords, identity cards, and biometrics. Identity Management (IDM) technologies are developed so as to manage the wealth of (personal) information circulating. These technologies control the access and restrictions of individuals (and groups) and manage their rights, entitlements, opportunities, and accountabilities.

IDM technologies may regulate the access to (physical) spaces, knowledge, information, communication infrastructures and a wide variety of services. They do so by means of granting authorization on the basis of verified, authenticated identities. In their most elaborate form, IDM technologies encompass all means necessary in an organization or infrastructure for electronic data processing regarding persons, including such issues as the registration of user names and preferences, (re-)identification of persons and the regulation of access to services and applications. In this broader sense, IDM technologies also facilitate the storage, processing, and use of elaborate digital files on persons in databases for a wider range of purposes than merely for the purpose of regulating access. In this sense, IDM technologies are closely connected with profiling technologies. Profiling technologies can be used to process this information into (group) profiles by means of statistical analysis and data mining, and structure (organizational) communication practices accordingly. From now on I will use the notion of IDM and profiling technologies to refer to this cluster of technologies.

People are increasingly sorted and classified by means of IDM and profiling technologies, for example for marketing, administrative or security purposes. This is done on the basis of the available data regarding their personal properties, such as purchase or travel behavior, income characteristics, postal code, and medical information. In addition, profiling technologies and statistical studies can be used to complement this information and get to know still more about a person. In this sense, profiling has an inherent epistemic aim regarding persons and their identities. The way they are recorded and represented in databases is therefore crucial for the treatment they receive with respect to their opportunities and rights. It leads to advantages and disadvantages for the individual whose identity is established. As a consequence, it seems as though the individual is burdened with another kind of *personal* 'identity management,' namely the management of one's 'identities' and profiles, i.e. one's personal, or identity related, information in relation to different purposes and different contexts. Examples of practical consequences as a result of the way in which

persons are registered or profiled, are the difficulties scientists with an Arab surname encountered when trying to enter the United States in the years following 9-11 (see Guardian: 2005). Another striking example is the case of the 'Burakumin' in Japan (see Alabaster: 2009):[4] prejudice and discrimination against these lowest-caste Japanese (reaching back to the Shogun feudal era) was reinforced when Google put up historical maps of Tokyo. Through the availability of these maps the areas where the Burakumin lived (and supposedly still do) could easily be identified, and used for example for screening out Burakumin-linked job seekers. Redlining practices by financial institutions likewise convey limiting effects for being granted a loan or mortgage, for example on the basis of postal codes.[5] For particular individuals the consequences pertaining to the way in which they are profiled or identified may come across as unjust. They feel they should be given a chance or opportunity in spite of the categorization, whether the categorization is accurate or not. They feel there is 'more' to them than the way in which they are known or characterized.

This thesis investigates the validity of this intuitive resistance from a moral perspective on identity. It explores the conceptual opposition between two conceptions of identity: On the one hand there is 'identity' as it is conceived of in moral discourse and often referred to in everyday life. I call this "moral identity." On the other hand, there is the notion of 'identity' as it is understood in a more practical, administrative sense, pertaining to the way it is often implied in IT discourse, what I call "practical identity." The overall aim, then, of this research, is to identify moral considerations to be used for informing the design of IDM and profiling technologies. This should a) contribute to a better understanding of what is at stake from a moral perspective concerning 'identity' in the context of IDM and profiling technologies, and b) provide designers with grips for better attuning these identity related technologies to the environment of their potential data subjects or users. This is what I call "Designing for Moral Identity."

It should be noted that this research proceeds from what some may refer to as an individualist conception of identity. It implies that the (moral) identity of persons consists primarily of a unique constellation of features and

4 After the effects of posting these maps became clear, they were removed by Google.

5 In the Netherlands there is a discussion whether the use of postal codes as a key determining factor for granting mortgages or insurance can be interpreted as risk limitation or discrimination, see for example http://www.discriminatiezaken.nl/doc/Factsheet%20postcodediscriminatie.pdf. (in Dutch)

characteristics, some of which are relational. Yet it does not imply that this conception entails an absolute right of individuals to be the sole author of their practical identities. And I advocate the reconciliation of this individualist conception of identity with the exigencies of the technological and societal structures in which we live.

Furthermore, the focus of this research is primarily on non-obvious, implicit moral risks concerning individuals and their (represented, practical) identities or profiles. This focus forms an important addition to the more obvious risks pertaining to this context such as identity theft, information-based harm, informational inequality, and (data) surveillance. These issues have received discussion in the literature. However, I think the issue of incongruence between moral identity perception and the way the concept of identity is implied in technological structures is a distinct moral and philosophical topic that deserves attention. So far, this topic has hardly been studied by philosophers. This study aims to fill this gap.

Especially in a context which has been characterized by some as a "risk society,"[6] decisions are increasingly made on the basis of probabilistic calculations. Risk discourse is firmly associated with modern technology: the risk of storing nuclear waste, of producing genetically engineered crops and of releasing nano-particles in the environment. In these cases, risk is usually predicated of events and construed as danger times the chance of the occurrence of events. Persons however, can also be characterized in terms of probabilities: probabilities that they will commit crimes, will like certain commercial products, are prone to accidents, are likely to exhibit certain types of unhealthy behavior, constitute moral hazards for insurance companies. The use of computer supported modeling techniques, computerized databases and statistical methods in fields such as law enforcement, forensic science, policing, taxation, preventive medicine, insurance, and marketing greatly promote the construal of persons as risks. I argue that if this probabilistic construal of persons and their characterization in terms of (risk) calculations were to become the dominant

6 This term is strongly associated with modernists Anthony Giddens and Ulrich Beck, referring to societies increasingly preoccupied with the future and associated uncertainties and risks. See e.g. Giddens, A. (1999) 'Risk and Responsibility', in: *Modern Law Review* 62 (1): 1-10, and Giddens, A. (1999) Runaway World: How Globalization is Reshaping Our Lives, London: Profile; Beck, U. (1992) Risk Society: Towards a New Modernity, Sage, New Delhi. For a comprehensive history of man's efforts to understand risk and probability, see e.g. Bernstein, P.L. (1998) Against the Gods: The Remarkable Story of Risk, Wiley.

view, it would come at the expense of a moral conception of identity. People might not get a fair chance to develop their moral identities as they aspire, like Vincent in the introductory example. They may end up needlessly fighting their own profile. For this reason I advocate an attitude of respect for persons with regard to epistemic claims concerning their identities.

Notably, I do not wish to discredit the gains and benefits as a result of the widespread application of IDM and profiling technologies. It is clear that these technologies have significantly improved (general) epistemic practices, enhanced the efficiency and scope of services, and contributed to the acceleration of organizational processes on a global level. Yet I claim that the ubiquity of these technologies demands the completion of their design with moral considerations concerning identity (perception), so as to contribute to the ongoing development of a just society.

This thesis addresses the concerns as described above. It aims specifically at formulating normative considerations for the design of IDM and profiling technologies. The research is embedded in the larger research program "Values in Engineering and Society" of Delft University of Technology and the 3TU Centre for Ethics and Technology. This program builds on a conception of doing ethics of information technology called "Value-Sensitive Design", that

> [A]ims at making moral values part of technological design, research and development. It assumes that human values, norms, moral considerations can be imparted to the things we make and use and it construes information technology (and other technologies for that matter) as a formidable force which can be used to make the world a better place, especially when we take the trouble of reflecting on its ethical aspects in advance. (Van den Hoven: 2007, 67)

As part of the research for this thesis, the approach of Value-Sensitive Design (VSD) is critically examined (Ch 5 – 6).

The first part of this thesis (Ch 2 – 4) consists of conceptual philosophical analysis on the basis of literature research. Its focus is on the 'context of justification' of values, norms, and moral considerations by examining some of the central moral concepts in relation to identity and IT. The context of justification pertains to the social, cultural, and moral milieu in which the design and use of a particular technology is justified. This milieu holds (implicit) assumptions regarding the design and use of the technology, and ethical analysis can be used to explicate and assess these presumptions. This thesis offers the

necessary preliminary conceptual work in order to substantiate moral claims and considerations regarding identity in relation to IDM and profiling technologies. These considerations can in turn be used for developing the actual technical requirements for a "Value-Sensitive," or "Value-Conscious," (cf. Ch 5) Design.

This thesis consists of 5 published articles that form chapters 2-6. In the remainder of this introductory chapter, I will first outline the thread that runs through this thesis, and highlight the key elements in light of the principal objective of this research. I will conclude with a brief outline of each of the enclosed chapters.

1.1. Designing for moral identity in information technology

Information technology (IT) presents us with a wide variety of means and options for constructing and expressing our (moral) identities. Over the past decade the Internet has evolved into what is called Web 2.0, promoting personal creativity and allowing users to actively engage in recognizing, identifying, constructing, and presenting oneself. Consider the recent emergence of social-networking sites, video-sharing sites, 'wiki's,' 'blogs,' and Twitter, to name only a few. At the same time, information technologies such as IDM and profiling technologies are used to structure and manage the wealth of identity related data associated with the ubiquity and advance of IT in everyday life.

These converging developments present a paradoxical situation with regard to moral identity development and the way in which we are increasingly categorized: whereas for one thing self-expression and self-development seem to be prominently promoted by means of IT, for another the advance of the use and application of IT seems to impose a forensic logic on account of which the identities of persons seem determinate, fixed, essentially quantifiable and manageable. The ubiquity of (probabilistic) profiles and categorizations by means of IT in this sense *limits* the extent to which we are able to freely construct our moral identities. I have characterized this paradox as the conceptual opposition between "moral" and "practical" identities (Ch 2).

The confinement imposed on the free and explorative development of our moral identities, is the result of a seemingly innocent process. This process consists of the collection and processing of excessive amounts of data. In some cases this data or information is uniquely referring, and collected or stored in files under the heading of particular labels or identities; in other cases the information is processed, e.g. subjected to data mining, resulting in new or

additional information regarding persons and their properties, or their profiles. Only when a certain threshold of information regarding one individual is accumulated in one way or the other, it holds the potential for "accumulative informational harm" (Ch 3), i.e. (moral) harm grounded on the accumulated collection of this data. Up to the threshold, the collection of each single piece of data often seems harmless. That is why the cause of accumulative informational harm is non-obvious and nearly imperceptible.

In addition, information regarding persons cannot always be clearly reduced to its subject. Descriptive phrases regarding persons can be used both referentially and attributively (Ch 4). Consider the phrase: 'The owner of a blue Ford.' This phrase can be referring to a specific person, in which case the descriptive phrase is used referentially; but it may also be a phrase about 'whoever fits the description,' in which case the description is used attributively. For individuals this implies that they can be described and characterized in both ways. That is why I claim (with my co-author in Ch 4) that the class of information about individuals that is worthy of protection does not only consist of personal information, i.e. according to its referential reading. Instead I propose to extend this category to what I call "identity related information" (Ch 4), so as to include attributively used descriptions that may turn out to refer to a particular subject in a later stage.

A way to approach the protection of identity related data so that the individual who the information is about remains involved, is to consider "moral identification" (Ch 4) as one of the basic principles for design. This notion involves epistemic modesty and respect for persons as self-presenters. It starts from the recognition of the first-person - or self-informative (Ch 2) - perspective of individuals with regard to their identities. Though still on an abstract, conceptual level, this notion can be used to inform the value-conscious design of IDM and profiling technologies. For example, this principle can be interpreted by means of designing tools specifically aimed at enabling flexibility for, and endorsement of, the individual who the information is about.

In order to develop a sound, practicable methodology for implementing moral values into design, there is still work that needs to be done (Ch 5). Recently various frameworks have been developed drawing attention to the incorporation of values into design, with "Value-Sensitive Design" as the most prominent and reviewed approach. However, taking values into account is not the same as normative reasoning. In order to have technology consciously and deliberately designed to include ethical value considerations, what I refer to as

"Value-Conscious Design," it needs to be complemented with the explication of normative objectives and the ethical theory it propounds to use, as well as the practical objectives for each particular design. In this way competing values can be confronted and trade-offs can justifiably be made. One of the suggestions to make such a methodology ready for use is to include a "values advocate" on design teams, i.e. someone who is in charge of mapping the ethical points of attention, stakes and aspects (Ch 6). This person is to explicate the normative objectives as well as the ethical framework to be used. By providing insight to the complexity and delicacy of value choices basing on theoretical knowledge as well as acquired practical expertise, the values advocate is able to educate the other members of the design team (and possibly other stakeholders) and to strongly promote certain value choices over others where necessary.

To conclude, this thesis provides the conceptual groundwork in order to arrive at a better understanding of what the concept of 'identity' entails with regard to the context of IDM and profiling technologies. This research can be used to inform the design of these technologies, so as to overcome the conceptual opposition between moral and practical identity perception. The considerations as suggested in this thesis comprise, amongst others, epistemic modesty regarding persons and respect for persons as self-presenters. In a more practical sense, this entails offering opportunities for individuals to endorse their own identities or profiles, and flexibility in design with regard to (value) preferences. In addition, I advocate limiting the collection, processing and use of information to clear objectives (Ch 3). However, in order to complete the aim of designing for moral identity in a practical sense, i.e. the formulation of practical requirements for (IT) designers in order to arrive at the value-conscious design of one of IDM and profiling technologies, still more work needs to be done. This not only comprises the translation of moral considerations into practical requirements, but also the further details of a methodology for implementing moral values into design. Finally, I hope I have successfully outlined the moral relevance of the conceptual field pertaining to 'identity' in relation to IDM and profiling technologies. Though far from complete, this research provides a start for addressing the implicit moral concerns arising from the ubiquitous and widespread identification of individuals on the basis of their acquired identities and profiles.

1.2. Chapter summaries

Chapter 2: Practical versus Moral Identities in Identity Management

Published in: *Ethics and Information Technology* 12 (1) 2010.

In this chapter, the notions of "moral" and "practical" identity are introduced. Whereas the term 'Identity Management' is used primarily in policy and technology settings to refer to ways and methods of dealing with registration and authorization issues regarding persons in organizational and service-oriented domains, there is another meaning of the term 'identity management,' which is clearly related and which has gained currency. This second construal refers to the need to manage our 'moral' identities and our identity related information. The chapter explores the relation between the management of our (moral) identities and 'Identity Management' as conceptualized in IT discourse.

I argue in this chapter that information technologies come with inscribed policies and criteria for (re-)identification. Information technologies impose a practical, i.e. an administrative, forensic, conception of identity and identification that is distinctly different from what a moral conception of self-presentation and identity management requires. This, I argue, unduly limits persons with respect to their self-presentation and management of their moral identity. Here I explore the conceptual opposition between these two different notions of identity and the fact that one of them is promoted by technology at the expense of the other. The chapter consists of an investigation of the context of identity and information technology, a discussion of the two different ways of construing the expression 'identity management' and associated conceptualizations of 'identity,' and three current phenomena demonstrating their conceptual opposition.

Chapter 3: Regulating Invisible Harms

Forthcoming in: Simone van der Hof and Marga Groothuis (eds.), *The Impact of Europe on eGovernment*, Information Technology and Law Series, Asser International Press, 2010.

This chapter describes the epistemic gains of IDM, as well as the type of harm that is central to this thesis. It discusses the benefits and harms of deploying IDM technologies for e-government. First, it is noted that IDM issues, though

rarely recognized, are connected to issues of justice. The way citizens are identified (if they are), has political or value implications for their opportunities in society. The balancing between individual and collective well-being is what is at stake in promoting fairness and equality. Therefore, identification is closely related to the design and management of a 'just' society.

The benefits of IDM technologies for one find expression in their epistemic gains. In comparison with traditionally available epistemic tools, I evaluate the success of IDM technologies in terms of generating and structuring valuable knowledge for (e-)government practices. There are at least five good epistemic reasons for adopting IDM, as I will demonstrate by applying Alvin Goldman's criteria for evaluating the epistemic success of social practices.

However, the 'informational' structuring of public administration and monitoring of citizens also implies certain risks, one type of which I further explore in this chapter. It concerns a type of harm which seems nearly invisible and negligible in each particular case, but becomes apparent in a collective, accumulated fashion. This, so I argue, also applies to the collecting and processing of personal - or identity related - data. It is what I call "accumulative informational harm," drawing on what Joel Feinberg has coined "accumulative harm."

Chapter 4: Identity Management and Moral Identification

This is a joint publication with Jeroen van den Hoven, submitted to *Science, Technology, and Human Values*. It is a translation of 'Identiteitsmanagement en Morele Identificatie', in: *Algemeen Nederlands Tijdschrift voor Wijsbegeerte*, 98 (2), 2006, 111-128.

This chapter develops one of the main moral requirements for 'Designing for Moral Identity,' i.e. the requirement of "moral identification."

As is well-known, persons are ubiquitously represented in databases. In this chapter I explore (with my co-author) the obligations we have to persons when we represent them in databases, when we record and use their identities, and when we interact with them on the basis of these representations. The chapter articulates the main moral reasons for imposing constraints upon the development and use of IDM technologies by persons other than those whose identities are at stake.

The chapter starts with a description of a type of identity, i.e. what I call "nominal identity" in Ch 2, and the associated rules of conduct (information

games). This is followed by describing two fallacies in current thought on personal data and the protection thereof in light of the potential offered by contemporary IDM technologies. Even without unique references and links to the observation of individuals, IDM technologies can be used to develop strategies for finding out more about people (whoever they may be). As a consequence, we propose extending the category of data concerning persons and their identities in merit of protection to what we call "identity-related" - instead of "personal" - data.

In the second half of this chapter, we discuss four moral reasons for protecting identity-related information in this broader sense. The fourth moral reason, pertaining to moral autonomy, calls for epistemic modesty with regard to the identifiability (knowability) of persons, and respect for persons as self-presenters. This leads to the formulation of the requirement of moral identification, i.e. the incorporation of the first-person perspective regarding persons and their identities, by drawing on Bernard Williams' use of the term.

Chapter 5: What Values in Design? The Challenge of Incorporating Moral Values into Design

Forthcoming in: *Science and Engineering Ethics*, online first at http://www.springerlink.com/content/d28305n23v9715t3/

In this chapter I explore if and how moral value requirements, such as the requirement of moral identification (Ch 4) can be designed into information technology. Recently, there is increased attention to the integration of moral values into the conception, design, and development of emerging IT. The most reviewed approach for this purpose in ethics and technology so far is Value-Sensitive Design (VSD).

This chapter considers VSD as the prime candidate for implementing normative considerations into design. Its methodology is considered from a conceptual, analytical, normative perspective. The focus here is on the suitability of VSD for integrating moral values into the design of technologies in a way that joins in with an analytical perspective on ethics of technology. Despite its promising character, it turns out that VSD falls short in several respects: (1) VSD does not have a clear methodology for identifying stakeholders, (2) the integration of empirical methods with conceptual research within the methodology of VSD is obscure, (3) VSD runs the risk of committing the naturalistic fallacy when using empirical knowledge for implementing values in

design, (4) the concept of values, as well as their realization, is left undetermined and (5) VSD lacks a complimentary or explicit ethical theory for dealing with value trade-offs.

For the normative evaluation of a technology, I claim that an explicit and justified ethical starting point or principle is required. Moreover, explicit attention should be given to the value aims and assumptions of a particular design. The criteria of adequacy for such an approach or methodology follow from the evaluation of VSD as the prime candidate for implementing moral values in design.

Chapter 6: Values And Pragmatic Action: The Challenges Of Engagement With Technical Communities In Support Of Value-Conscious Design

This is a joint publication with Michael Zimmer and is forthcoming in Emma Felton, Suzi Vaughan, Oksana Zelenko (eds.), *Ethics and Design*, Sense, 2010.

This chapter focuses on the pragmatic challenges of engaging in the implementation of moral values in design, especially in real-world design settings.

Recently, focus has been placed on how to develop pragmatic frameworks that ensure that particular attention to moral values becomes an integral part of the conception, design, and development of emerging information systems. Each of these frameworks – which I and my co-author refer to collectively as Values In Design – seek to broaden the criteria for judging the quality of technological systems to include the advancement of (human and) moral values, and to proactively influence the design of technologies to account for such values during the conception and design process.

In this chapter two attempts to practically engage with (technical) design communities in order to influence the design of these technologies in value-conscious ways are described, revealing discouraging results. Learning from these failed attempts, this chapter identifies three key challenges of pragmatic engagement with technical design communities: (1) confronting competing values, (2) identifying the role of the values advocate, and (3) the justification of a value framework. Addressing these challenges must become a priority if one is to be successful in pragmatically engaging with real-world design contexts to support the value-conscious design of emerging information technologies. To help illuminate how these challenges might play out in real-world design scenarios, we have contextualized them in relation to a possible future

application of Values In Design to address the privacy implications of Web search engine query data retention.

1.3. References

Alabaster, J. (2009) Google crosses line with controversial old Tokyo maps, Retrieved January 11, 2010 from http://search.japantimes.co.jp/cgi-bin/nn20090505a1.html.

Guardian. (2005) Retrieved January 11, 2010 from http://www.guardian.co.uk/education/2005/jul/15/highereducation.uk.

Lundry, A. (2004) 'Microtargeting: Knowing the Voter Intimately,' in: *Winning Campaigns Magazine* 4 (1).

Van den Hoven, J. (2007) 'ICT and Value Sensitive Design', in: Goujon, P., Lavelle, S., Duquenoy, P., Kimppa, K., Laurent, V. (eds.), *The Information Society: Innovations, Legitimacy, Ethics and Democracy*, Springer, Boston: 67-72.

2 Practical versus Moral Identities in Identity Management

This chapter appeared as an article in *Ethics and Information Technology* 12 (1) 2010.

2.1. Introduction

Information technology is becoming ever more pervasive in our everyday activities such as communication (both private, public, and in work), the use of business and government-related services, and the market economy. In each of these areas, information about us is collected and stored in databases for multiple purposes. Decisions that may affect us deeply, e.g. regarding our entitlements, our opportunities, and even our personal environment (as in the case of ambient intelligence), are made on the basis of personal profiles and available (identity related) data. So information technology facilitates better adjustment of products and services to our particular needs and preferences, but it also makes available a wealth of information about us, including associated profiles and reputation records.

Over the past decade Identity Management has become a central theme in information technology (design), policy, and administration in both the public and private sector.[7] In these contexts the term 'Identity Management' is used primarily to refer to ways of dealing with registration and authorization issues and access of persons in organizational and service-oriented domains. Especially due to the growing range of choices and options for, and the enhanced autonomy and rights of, employees, citizens, and customers, there is a growing demand for systems that enable the regulation of rights, duties, responsibilities,

[7] The idea of Identity Management was first conceived by David Chaum, (1984) who proposed a card-computer for each individual to handle all transactions. The term 'Identity Management' was first brought to our attention by several consultant companies: "The name Identity Management actually appeared around the year 2000, and was cited as such by important consultancy companies such as the Burton Group and Gartner. It had become a common denominator for a variety of technologies focusing on information security." Frank Ramdoelare Tewari: 2005, 25.

entitlements and access of innumerable people simultaneously. 'Identity Management' or 'Identity Management Systems'[8] are now important rubrics under which such systems are designed and implemented. But there is another meaning of the term 'identity management,' which is clearly related and which has gained currency. This second construal refers to the desire, need and right to manage our (moral) identities – and thus our identity related information - in the context of different information technologies. This paper articulates the conceptions of identity that are presupposed by the use of the term 'identity management' in these two senses. A clear understanding of the implied notions of identity can be used to facilitate new and better design strategies of IT applications in the area of identity management which take into account moral considerations concerning identity in a morally relevant sense.

Information technology is a highly efficient and effective way of organizing, gathering and processing information in a broad range of contexts. The number of possibilities and the reach of monitoring information technologies exceed human capacities by far (see Ayres: 2007). For this reason new IT applications and tools are constantly being developed which enhance our limited capacities for information processing.[9]

Over the past decade the Internet has evolved into what is called Web 2.0, allowing users to do more online than just retrieve information. Information technology, especially as a central means of communication, offers a wide spectrum of opportunities for recognizing, identifying, constructing, and presenting oneself; this is done for example through applications such as social-networking sites, video-sharing sites, 'wiki's,' 'blogs,' and Twitter,[10] to name only a few. The Internet has been praised as an enhancement of the public forum and as a platform for democracy; whereas the threshold for engaging in (public) discussion as well as private communication is low, the reach is potentially

---

8 Please note, 'Identity Management' in capitals refers only to the way it is used in IT discourse.

9 Consider for example governmental matters such as free-rider problems in relation to the distribution of public goods, business cases involving trade representatives and delegates, or classic theoretical examples such as the prisoner's dilemma. The rationality of the situation in these cases calls for the gathering and processing of information to reduce information asymmetry between parties and that is exactly why the deployment of information technologies is not a passing phase, but here to stay.

10 Twitter is a recent IT application that has swiftly gained popularity. Its aim is keeping others up-to-date by means of continuously updating information on one's whereabouts through (mobile) Internet. See http://twitter.com/

staggering.[11] It enables and promotes personal creativity and realization of modes of (self-)presentation.[12]

What I argue in this paper is that information technologies come with, or have inscribed in them, policies and criteria for (re-)identification. By implication, they impose an administrative, or forensic, conception of identity and identification that is distinctly different from what a moral conception of self-presentation and identity management requires and they may thus unduly limit persons with respect to their self-presentation and management of their moral identity. This paper explores the tension between these two different conceptions of identity and the fact that one of them is promoted by technology at the expense of the other.

I start with an investigation of the context in which this tension is situated. This is followed by the introduction of two different types of identity that are relevant to understanding the notion of moral identity. Next, two different ways of construing the expression 'identity management' are discussed, i.e. in the context of IT and in the context of moral discourse. Then I discuss identity construction in a moral sense in relation to normative contextual expectations. Finally I discuss three phenomena which demonstrate the clash between the two different senses of 'identity management.' To conclude, I suggest how this analysis contributes to a better understanding of what is at stake in modern identity related information and communication technologies and how this understanding may lead to a better design of information technologies.

2.2. Super crunchers

Before I further explore possible conceptualizations of identity, identity management, identity construction, and the way this relates to the context of information technology, this context and the outlook it facilitates is described.

As a result of the growing need for gathering and processing of information and the associated rapid development of new information technologies, the world is seemingly becoming more quantitatively oriented and much of our decision-making is increasingly informed by probability calculations and risk

[11] Yet in his book *Republic 2.0*, (2007) Cass Sunstein warns against the opportunist use and filtering of information on the Internet, endangering the core values underlying democracy.

[12] Even before, during what may be called Web 1.0, people were already encouraged to present themselves e.g. by building personal websites or connecting to peers via different (news) groups and emailing lists.

assessments. This is part of what Ian Hacking has referred to as the "probabilistic revolution" (see Hacking: 1987, Gigerenzer *et al.*: 1989).

In an optimistic account Ian Ayres (2007) points to the previously unknown and tremendous impact and opportunities for prediction as a result of large-scale data analysis by "super crunchers," as he calls data analysis technologies in his book with the same title. According to Ayres, data mining not only completely supersedes human capacities for data processing, but it also seems to make human predictions and decision-making obsolete. To put it stronger: human expertise is in many contexts superseded by the capacities, promises, and widespread application of algorithms.[13] In his introduction, Ayres relates the story of Orley Ashenfelter, a wine-loving economist at Princeton who wanted to safeguard and optimize his future wine investments, and developed a regression model for determining what factors are important for predicting the quality of future Bordeaux wines.[14] The model received strong critique from the wine expert community at first (most of its members strongly committed to the conviction that the quality of wine can only be determined after a wine is produced), but eventually proved to be quite successful in its predictions. Ian Ayres points to a shift in science and policy, where predictive models and analytics are winning ground.[15] The financial world and stock market are a cradle of constant innovation of tools and applied mathematics in the service of predicting the 'futures.' This way of thinking and proceeding spills over into other domains of life so that today members of Olympic teams are picked on the basis of econometric analyses of their track records.

I claim that this shift furthers thinking in terms of a paradigm in which everything is made to look determinate, tangible and essentially quantifiable: while the success of the application of information technology to issues where the outcome can be expressed in terms of (numerous) parameters seems obvious, it is questionable whether this is also the case regarding human beings

[13] However, in his book *Gut Feelings: The Intelligence of the Unconscious* Gerd Gigerenzer (2007) provides contrary evidence.

[14] The factors turned out to be winter rainfall, average temperature and harvest rainfall.

[15] In 'The Automation of Discovery' (2004) Clark Glymour also proclaims a paradigm shift for science, where the traditional model of hypotheses being tested by research is overthrown by the investigation of random patterns in multiple linked databases (data mining), turning attention in academic research to explore the deviant patterns of statistical mining instead. For a discussion of the difference between data mining and statistics see for example D.J. Hand (1998) 'Data Mining: Statistics and More?.' On his account, data mining is different in that it does not require a particular hypothesis before analysis.

and their identities and personal properties. Can the identity of a human being be expressed as, or captured in terms of data sets? Or is a conceptual gap bound to remain between identity in information technology and identity in moral discourse?

The application of statistical reasoning to social phenomena was first introduced by Adolphe Quetelet in the 19th century. Quetelet believed that the coherence of social phenomena could be demonstrated by means of statistically rendering these phenomena.[16] He showed a strong confidence in statistics and the associated development and applicability of (social) laws. He developed a statistics of deviance, which he coined "analyse morale," referring to the classification of people into categories developed by means of statistical analyses, and the possibility of characterizing them as 'deviant' to an objective mean accordingly (see Quetelet: 1842). As a result, human variation, social, behavioral as well as physical, was interpreted as a variation from an objective mean, distributed according to a law of errors (see Gates: 1992). This marks an important turn in the way human beings and their identities were perceived, namely in relation to each other: No longer were humans defined according only to themselves, but now also in relation to each other and in relation to and as variations from Quetelet's new creature, *l'homme moyen*, The Mean or Average Man (see Coven: 2003; Gates: 1992). Presently, we find ourselves regularly defined in relation to others in IT-mediated environments.[17] This happens for example when we are seeking information on the Internet and using Google as a search engine: The search behavior of all users is recorded and analyzed in order to present the searcher with the most relevant search results. And when we buy a book at Amazon.com, we receive information on the preferences of other customers: "someone who bought this book also bought X..." For processing technologies, the so-called *super crunchers* – consider 'preference engines,' 'collaborative filtering,' and 'recommendation tools' to name a few-, we appear as statistical objects of study, abstracted from our personal preferences and life plans, and from our individual capacities and freedom to choose.

---

16 Victoria Coven, 2003. During and prior to this period a commonly held belief was that the occurrence of these phenomena was a demonstration of God's ordering of the Universe.

17 Information technology, especially by means of the Internet, makes it easy to collect and process large quantities of information on the preferences and actions of its users. What happens is that our behavior and preferences are tracked and recorded, e.g. by means of cookies, stored in databases and processed in order to gather insight into the variation and distribution of human action.

This, so I argue in this paper, conflicts with the way we most often perceive ourselves, with what I shall call our "moral identities." In certain contexts the imposition of a quantitative or numerical outlook inscribed in the structure of information technology gives rise to a conflict between competing conceptions of 'identity' and its management, which explore in more detail below. In some cases this tension turns out to be morally problematic.

2.3. Self-informative identity and nominal identity

Erving Goffman discusses three contending types of identity in his book *Stigma*, notably with an apt reference to 'identity management' in its subtitle *Notes on the Management of Spoiled Identity*. These three types of identity are referred to as "personal identity," referring to the uniqueness of one person, "social identity," referring to the (social) attributes of a person, and "ego identity." [18, 19] He suggests that social and personal identity are in some sense opposed to the latter:

> Social and personal identity are part, first of all, of other persons' concerns and definitions regarding the individual whose identity is in question. In the case of personal identity, these concerns and definitions can arise even before he is born and continue after he has been buried [...] [E]go identity is first of all a subjective, reflexive matter that necessarily must be felt by the individual whose identity is at issue. (Goffman: 1963, 129)

Thus, whereas constructing an "ego-identity" is primarily reflexive, the construction of a "personal" or "social" identity is primarily attributed by others, and proceeds by means of what I call "attributed identification." Goffman (1963, 130) then continues that although these different types of identities may be constructed out of the same building material(s), the individual exercises important liberties as to what he fashions for his (ego-)identity construction. One of the central liberties that Goffman suggests in this respect concerns the feelings of the subject relating to his or her ego-identity. Elsewhere in this issue, David Shoemaker (2010) argues similarly. He follows David Copp[20] on the importance of self-esteem to our self-conception: According to this view, what

[18] Here Erving Goffman draws on Erik Erikson's use of the term, amongst others. See Erik Erikson (1968) *Identity: Youth and Crisis*, Norton, New York.

[19] "Ego-identity" covers more or less the same idea as what I call "self-identity," whereas "personal" and "social" identity correspond with what I call "nominal identity" later on.

[20] For a discussion of what he calls self-esteem identity, see David Copp, 2002.

constitutes our moral identity is determined by what we value and respond to emotionally in relation to our self-esteem, be it pride, fear, shame.

There is another way in which "attributed identification," with or without the help of technology, differs from self-identification or constructing an identity of oneself. For this I will be drawing on the semantics of self-knowledge by John Perry (forthcoming). Let us assume that a constructed identity, whether reflexive in the case of self-identification or attributed in the case of identification by others, is made up of propositions. So, if we were to open up a mental file on say, Barack Obama, we may start filling this file with all sorts of beliefs we have about this person, e.g. 'Barack Obama is a man,' 'Barack Obama is the President of the United States,' 'Barack Obama lives in Washington.' These beliefs may be true or false. We could for example believe that 'Barack Obama has ancestors in Nigeria' when in fact he has ancestors in Kenya. Whether propositions are true depends on whether they meet certain truth conditions, a semantic discussion which I will put aside for the purpose of this paper. What is relevant though is that these truth conditions, following John Perry, incorporate the way a reference is achieved.[21] In other words, what is relevant is the source of knowledge regarding the constituents of the propositions: Who is this information about, and how is it obtained? So even if Barack Obama utters the two following propositions with the same referential content (1) 'Barack Obama lives in Washington' and (2) 'I live in Washington' and these seem to have the same cognitive significance, there is still an important and relevant difference on the level of content. In the latter case, the knowledge or content of this proposition is from a first-person perspective, or what John Perry calls "self-informative knowledge." The constituent of this propositional belief is Obama's 'self-notion,' or 'I.' [22] This can be distinguished from the first utterance made by Barack Obama: 'Barack Obama lives in Washington,' which is an utterance or proposition about 'the person he happens to be' (see Perry: forthcoming). In this case we say the proposition is about a *nominal* notion he has of himself, attached to a property of himself, such as his name. The key point here is that self-beliefs (de se) have a reflexive condition of truth *in virtue of the identity relationship one has to oneself*, as opposed to the non-reflexive and conventional condition of truth

[21] John Perry: forthcoming. There is a difference between 'referential contents' and 'conditions of truth.' The latter keep track of the roles objects must play, or the properties they must have, in order to be referred to. Cf. the well-known example of 'the present king of France.'

[22] John Perry claims (forthcoming): "We all use these same methods to find out things about ourselves, but can't use them to find out about others."

associated with someone's name, registration number, or any other identifier. To conclude this semantic excursion, the propositional knowledge of attributed identification has different conditions of truth than first-person, self-informative knowledge about one's self-identity. By implication, anyone can have nominal knowledge about an individual's identity, whereas self-informative identity knowledge is possible only in case of an identity relation to oneself. This marks the distinction between what I call *self-informative identity* or self-identification, and *nominal identity*, resulting from attributed identification.

Up to now I have discussed two ways in which self-informative identity differs from nominal identity: (1) the presence or absence of feelings regarding one's (self-) identity, and (2) the truth conditions related to the semantic content or propositional identity knowledge in case of self-beliefs versus nominal beliefs. There is I think a third distinction that sets the two apart and which can be brought out by introducing George Herbert Mead's distinction between "play" and "game" (see Mead: 1934, 151). Mead argued that children begin building a self-informative identity by (role-) playing, making up rules as they go along and at liberty to take on any role as they like. Games however, are more constrained, with constitutive rules determined by others.[23] In the latter case children are required to position themselves as 'selves' in the roles as defined in the game, in contrast with the flexibility when playing, for taking on roles of (distinct) others and continuously reshaping oneself. This ties in with the distinction suggested by Goffman between social and personal identity on the one hand and ego identity on the other. Whereas social and personal identity are initially attributed, i.e. in accordance with the rules and (normative) expectations of the other players one is engaged with (cf. "game"), ego identity on the other hand entails the liberty to shape and manage one's self-informative identity (cf. "play"). The way we are characterized, understood, and identified from an attributed or objectified perspective implies social organization and structure (of which we are part ourselves), while the experimental, reflexive construction of our self-informative identities is more fluid, changeable, unconventional, and associated with feelings. The discrepancy between self-identification and attributed

[23] Mead calls this the child's first encounter with "the generalized other," which is one of the main concepts Mead proposes for understanding the emergence of the (social) self in human beings. The generalized other can be seen as understanding the given activity and the actor's place within the activity from the perspective of all the others engaged in the activity. Through understanding the generalized other the individual understands what kind of behavior is expected, appropriate and so on, in different social settings. The mechanism of perspective taking within social acts is the exchange of social positions.

identification then is about the freedom to experiment with and change one's (self-)identity versus the normative expectations brought about by the constitutive rules in a social or technological structure.

Up to this point I have distinguished two perspectives on identity and identity construction. The first is (constructing a) self-informative identity, associated with self-beliefs, the freedom to experiment with and change one's identity (as in playing), and the importance of evaluating one's (self-)identity emotionally. The second type of identity, which I call "nominal identity," is constructed in relation to, and in part by, others. It is constituted in accordance with the constitutive rules set by others (as in games), such as those inscribed in the technological structure of information technology. Moreover, the beliefs associated with the construction of a nominal identity are (about) an attribute or *property* of the identity subject, for instance a name, label, or identification number. Self-informative beliefs on the other hand have a reflexive (identity-)relation to the subject the beliefs are about. For this reason the truth conditions of the propositions or beliefs regarding identity from an external or third person perspective are non-reflexive and conventional (nominal identity), as opposed to the reflexive and self-informative conditions of truth regarding self-beliefs (self-informative identity). Nominal identity as it figures prominently in administrative processes and forensic investigations entails a third-person, attributed perspective, whether from the person whose identity is at stake or from others. In this paper I explore the way in which these two types of identity relate to each other in the context of Identity Management and profiling technologies - or "super crunchers"- and how they relate to the two different notions of identity management.

2.4. Identity Management

In the previous section, I have distinguished two notions of identity: self-informative identity and nominal identity. Here I consider the two ways of construing the expression 'identity management.' As I will show, 'identity management' in moral discourse involves managing both types of identity, i.e. self-informative and nominal, whereas 'Identity Management' in IT discourse presupposes a forensic logic of (re-)identification, based on a single and unified conception of identity abstracted from the distinction made above.

'Identity Management,' in the sense of an IT application or infrastructure for regulating access rights, presumes a conception of identity that is related to its

technological format and the aims and principles of its design. Its main focus is on sameness and continuity,[24] such as re-identification and classification: Is X the same person who has accessed the system before? Does this person possess certain characteristics on the basis of which he or she is granted certain rights? Can this person be profiled using available identity information? Inscribed in its structure, information technology imposes a forensic, administrative notion of identity such that users can be (re-)identified repeatedly and consistently. It is also important to note here that this forensic notion of identity promoted by (IT) IDM applications is transitive. It would be completely useless for its purposes if it would not support judgments of the type 'A is identical to B, and B is identical to C, therefore A is identical to C.'[25] This contrasts with the moral notion of identity, which supports only similarity relations and family resemblances and not necessarily transitivity. Consider the retrospective view of an old man: "I was a romantic fool at the age of 17, who later turned into a serious mathematics student. As a retired mathematics professor I smile when I read my romantic love poems, and I find it impossible to relate to the way I felt when I wrote them."

As touched upon before, identities as understood in IT discourse, e.g. in identification technologies[26] such as Identity Management and Profiling technologies, are considered principally as *determinate*: they consist of labels, properties, attributes or files made up of attributed identifications or nominal beliefs regarding a person. This way, the information can be processed for different purposes such as profiling. An identity in this sense is a (data) set of potentially innumerable propositions, parameters or properties; but a person's *self-informative identity* is typically left out.

Yet as pointed out above 'identity management' (see Goffman: 1956, 1963) can also be construed in another sense, namely in the way we manage our identities, both our nominal and self-informative identities, the way we present ourselves to others and to ourselves. Jan Bransen uses the notion 'identity management' to refer to the process of self-development:

---

[24] Several authors have connected this to Ricoeur's notion of *idem* identity, as opposed to *ipse* identity. See Mireille Hildebrandt (2006, 2008), Charles Raab (2009), Paul de Hert (2008), Katja de Vries (2010).

[25] Thanks to Jeroen van den Hoven for pointing this out.

[26] There is a distinction between IT principally aimed at (re-)identification, which may be referred to as "identification technologies," and IT in general, including for example Web 2.0 applications. The latter includes for example social networking sites, where the emphasis is (also) on self-presentation and identity construction.

> [L]iving a life can insightfully be understood as a process to be managed, a process that involves, at least from time to time, solving practical problems by making difficult choices and that is related to its identity as its goal or telos. (Bransen: 2008, 107)

In addition to 'identity management' as an expression to refer to the regulation of access, rights and duties of administrative identities by means of IT, the expression can thus also be understood as referring to the ordering and structuring of one's moral identity as a life project. Moral identity is to be understood here as a comprehensive notion of identity involving both nominal and self-informative identities. It is a rich, thick concept referring to the way an individual wants to live her life (in relation to others). Bransen underlines the aspect of self-informative identification. According to Bransen, a person is engaged in 'identity management' by choosing alternatives *of* oneself[27] with respect to one's life choices, and constituting a self accordingly. I would add 'identity management' in the moral sense also applies to the shaping and coordinating of our self-presentations, hence our nominal identities. The constitution of a 'moral identity' then is the overall, general objective of this 'identity management.'

In summary, 'Identity Management' in the IT sense presupposes a conception of identity that is fixed, determinate, and consists of attributed, i.e. nominal information only. 'Identity management' in the moral sense concerns both nominal and self-informative identities as part of one's moral identity. The management of these various identities, which can also be controversial or conflicting, is a continuous, dynamic, reflective process. Yet engaging in a society or structure demands the careful orchestration of these identities into one comprehensive, intelligible and coherent account of identity, as discussed below.

[27] An expression and idea developed by Jan Bransen on several occasions (1996, 2000) referring to the way certain (important) deliberations and choices in life have consequences for determining the role and qualifications of a person's identity.

2.5. The construction of a practical identity

As indicated, 'identity management' in moral discourse involves managing both our nominal and self-informative identities as part of one's moral identity.[28] In order to gain a clear understanding of what is morally at stake in 'identity management,' this section further explores the way in which such moral identity is established and perceived and what conditions it should meet.

There is a wide agreement that identity construction is an individual affair only to a certain extent. Many authors have argued that central to identity is its social embeddedness.[29] According to Mead, it is within and through a community that we experience ourselves as a (conceptually unified) self (see Mead: 1934, 90; 1962, 154-158). Creating a nominal identity, a narrative, or a third-person construction of the self as Paul Ricoeur (1992) has called it, is the process of generating an intelligible account of one's identity in light of one's history, acquired understanding, or discretion and prevailing values.[30] In order for an identity to be practically useful, it must meet the requirements of coherence and intelligibility (see Velleman: 2001; Korsgaard: 1989; Schechtman: 1996). Notwithstanding the countless alternative narratives of us, including fictitious and self-deceptive ones, in order to be and conceive of ourselves as practical agents, capable of making intelligible decisions, acting, carrying out a rational life plan, directing one's single body, and deliberate on conflicting desires, the one identity that we use to designate our self must provide unity.[31] Velleman (2001, 53) subscribes to this view, holding that "insofar as you want to be eligible for social intercourse, you must offer a coherent public image." And Goffman (1963, 92) observes that in spite of the multiplicity of 'identities'[32] that role and audience segregation allow a person, there is a (social) structure holding the individual to one biography. This question concerning the practicability of (moral) identity applies in the context of IT in a broad sense: it ties in for

---

[28] For the sake of brevity both are sometimes referred to as 'identity' in this section.

[29] For Mead, an experience of the self is only possible in relation to others: "no experience of a self simply by itself." (1934, 195) See also Charles Taylor in The Sources of Self: "a self only exists among other selves" (1989, 35), "one cannot be a self on one's own" (1989, 36). See also Goffman (1956, 1963), Rom Harre (1983), Howard Kamler (1994). MacIntyre even speaks of "that newly invented social institution: the individual." Alasdair MacIntyre: 1984, 228.

[30] Taylor would add - and against the background of one's normative sources, cf Taylor (1989), Korsgaard (1989).

[31] This is what Christine Korsgaard (1989) refers to as the "coherence" criterion.

[32] Goffman uses 'selves' in the text to refer to what I call 'identities' in this paper. Supra note 36.

example with what is described in this paper as the opportunities provided by information technology, e.g. in Web 2.0, for self-expression and self-creation on the one hand, versus the normative expectations embedded in the structure of information technologies, i.e. concerning (forensic) (re-)identification, on the other.

The (life) project of constituting and shaping a moral identity is an enterprise in relation to oneself and to others.[33,34] Ricoeur (1992, 116) depicts this process as the permanent search for an answer to the question of "Who am I?" It involves continuous interplay with others regarding one's attributed identifications, similarities with others, and the moral and emotional evaluation of possible (self-)identifications. According to Ricoeur, attributed or third-person constructions of the self are necessary in order to develop variations of ourselves, and subsequently shape our moral identities.[35] Information technology has become an increasingly important means for this, especially to younger generations. The way they conduct their social lives and express their societal engagement is often expressed by means of online activities in social networks, fora, and blogs. These activities have become important means for constructing and shaping a moral identity in relation to others. Here, and even more so in online gaming environments and virtual worlds, persons take on different 'identities'[36] for role-playing and experimenting, in some cases to escape reality constraints.[37] A question that comes up when a person is engaged for example

---

33 The self as relational concept is alluded to by many authors. John Perry (forthcoming) defines identity as a relation one has to oneself. And Soren Kierkegaard (1849) claims that relating oneself to oneself and to others is what constitutes the self.

34 Dean Cocking and Jeanette Kennett (1998) point to the importance of the recognition and interpretation of one's character traits (identity) for close relationships. In holding this view, Dean Cocking and Steve Matthews (2001) argue that relationships contribute to the development of one's identity in an importantly relational way, e.g. by coming to see the world (and oneself) through the eyes of another.

35 Occasionally I will use 'constituting' and 'constructing' a self- informative identity interchangeably. Although I will not go into this here, I believe the emphasis of 'constructing' an identity is more on the creation of and experimenting with a self-informative identity, whereas the emphasis of 'constituting' an identity is more on the presentation of oneself (to others).

36 In order to distinguish between self- informative identity and nominal identity as discussed before and 'identity' as a role or character a person chooses to develop and perform, I've put the latter between inverted commas.

37 Marya Schechtman (1996) discusses reality and articulation constraints to an identity-constituting self-narrative.

in *Second Life*[38] in the capacity of another 'identity'- or even as multiple 'identities,' is how to relate these 'identities' to his or her self-informative or nominal identity. In 'Bodies, Selves,' David Velleman (2008, 407-412) argues that the constitution of a self-informative identity[39] in a virtual world is no different than constituting a self- informative identity in one's own body. According to him, in virtual worlds we are not 'pretend-playing,' but literally performing actions. The actors "experience themselves *as* the characters, behaving *in* character, under the impetus of their own thoughts and feelings." The manager of these different 'identities' then is faced with the challenges of coordinating and in some cases integrating these different 'identities' into one identity that gives rise to moral reasons and obligations. This is what Christine Korsgaard has termed "practical identity."[40]

Practical identity as such is a fruitful conception for discussing the (conceptual and practical) ethical challenges for identity management in both senses of the expression. It comprises both the forensic conception of identity as presupposed in the IT-sense of 'Identity Management,' as well as the orchestration of multiple nominal and self-informative identities as a result of 'identity management' in the moral sense of the expression. Consider a person who is a modest and loving father, a harsh financial director of a multinational and a devoted church member (amongst others) at the same time. While the coordination and alternation of these roles clearly involves 'identity management,' this is done within a structure of normative requirements regarding one's practical identity. This being said, meeting the conditions of coherence and intelligibility requires a certain degree of self-understanding on behalf of the manager or agent.

For many authors this understanding is obtained by way of narratives (see Ricoeur: 1992; Taylor: 1989; MacIntyre: 1984; DeGrazia: 2005; Velleman: 2005). In the words of one of the most prominent advocates of the narrative account, Marya Schechtman, "[t]o have an autobiographical narrative in the

---

[38] Second Life is one of the main virtual worlds at this time, launched by Linden Lab in 2003. For more information see http://secondlife.com/.

[39] Here I use the term 'self- informative identity', whereas David Velleman speaks of "constituting a self" in the text.

[40] Christine Korsgaard: 1994, 83-84. I consider one identity to be the basic principle for understanding and taking part in this world. However, as pointed out by an anonymous reviewer, agential selfhood is a kind of achievement individuals sometimes struggle with in practice (as in the case of pathological dissociative disorders), and it is something that obtains in degrees.

relevant sense is [...] to have an implicit understanding of one's history as unfolding according to the logic of the story of a person's life" (1996, 113-114). The biographical and narrative interpretation of moral identity is the notion of identity that is often referred to in everyday life, when discussing one's self-understanding and relationships to others in the world. It covers one's life story; identity as embedded in a social context with memories, contacts, role-playing and emotions. For some it is the basic condition of making sense of ourselves.[41] Charles Taylor (1989, 47) even argues that the narrative form is "an inescapable structural requirement of human agency." He claims that a narrative understanding of one's life goes beyond merely structuring the present. It is about *who I have become*, analogous to orienting in (moral) space: "In order to have a sense of who we are, we have to have a notion of how we have become, and of where we are going." The narrative of oneself includes projections of a future story, a "sense of my life as having a direction towards what I am not yet" (see Taylor: 1989, 48). He compares this to Alasdair MacIntyre's notion of (the unity of) life as a "narrative quest" (see MacIntyre: 1984, 218). In keeping with this view, constructing a moral identity is defining oneself in relation to others in light of a continuous reflective process, evaluating and identifying with one's attributed identifications and personal choices in life and engaging in self-improvement.

Although the narrative view, stressing the importance of narratives for gaining understanding of our moral identities and lives, and even for providing direction to our lives (see Taylor: 1989; MacIntyre: 1984), is widely supported, some authors have questioned its limitations with respect to the constitution of our (self-informative) identities. David Shoemaker (2010), for example, doubts the extent to which narrative theories can (fully) accommodate attitudes and experiences, e.g. of physical characteristics. Moreover, he points to the subjective nature of such account. He proposes following David Copp instead,[42] focusing on the importance of self-esteem to our self-conception. On this view, our moral identities are constituted by properties or beliefs about ourselves we value and

[41] Charles Taylor: 1989, 47, Paul Ricoeur: 1992, 158. Galen Strawson in 'Against Narrativity' (2004) fervently disagrees with the unity and form-finding construction narrative theory proclaims. He points to the underlying ideals of control and self-awareness and the associated tendency for correspondence. He claims that the proponents of the narrativity-thesis impose their personal views on others whereas others may well experience their lives 'episodically', i.e. without a sense of diachronic unity of past, present and future.

[42] For a discussion of what he calls self-esteem identity, see David Copp (2002).

respond to emotionally in relation to our self-esteem.[43],[44] This view, still narrative in part (see Shoemaker: 2010), also comprises the social (hence the nominal) dimension of identity construction.

To recapitulate, the narrative account contributes to an improved understanding and the construction of our self-informative and nominal identities. However, for moral identity, i.e. both self-informative and nominal identity, to be practically useful it demands compliance with the constitutive rules of the "game," e.g. taking part in the information society or in social intercourse (see Mead: 1934). This is where both senses of 'identity management' meet: 'Identity Management' technologies have inscribed in their structure a practical, forensic, administrative notion of identity such that users can be (re)identified reliably, whereas 'identity management' in the moral sense of the expression involves the shaping and constructing of both nominal and self-informative identities prior to the integration into a practical identity.

2.6. Three points of tension

As said before, recent advances in information technology have enhanced the possibilities for presenting and molding identities online, e.g. by means of encouraged participation in Web 2.0 applications and opportunities generated as a result of the implementation of fuzzy logic[45] to create complex decision infrastructures. Contemporary developments in online behavior include the exchange of personal information, preferences, and experiences in what are called 'social networks,' and the possibilities of refining one's preference sets in both the public and private domain.[46] Especially with the combination of vision, sound and the increasingly ubiquitous dimension of 'connectedness' in

---

43 For Shoemaker (2010) this serves as an explanation of why we may object to data mining of seemingly trivial information without clear sight of the potential harm it may bring upon us: "I am [...] unable to be the manager of my own reputation. Breaches of informational privacy thus take away a key element of self-determination [...]."

44 This connects to Goffman's ideas mentioned earlier (p.40) concerning feelings as one of the central liberties of "ego-identity."

45 Fuzzy logic is a mathematical method of handling approximate rather than precise concepts and values. Other than classic propositional logic, which has only two possible truth values (i.e. 0 and 1), fuzzy logic is a multi-valued logic where the degree of truth of a statement can range between 0 and 1.

46 An example of this would be the active participation of a citizen in (local) governmental decision-making by means of the Internet or someone indicating one's notification preferences with respect to administrative services online.

information technology, a world of possibilities has opened up to the construction and enacting of our moral identities. These identities may vary infinitely in arrangement and tenor. Concurrently information technology is used as a means for categorization, registration, and identification. This is what often comes under the header of Identity Management and Profiling.[47] In this section I discuss three phenomena associated with identity construction in the context of information technology that bring out the tension between 'Identity Management' as it is commonly understood in (policy and in) IT discourse and 'identity management' in the sense of being engaged in constructing our self-informative and nominal identities.

But first, let me make clear what is *not* at stake here. As is widely known, data processing brings with it the omnipresent technical risks of information technology such as: viruses, hacking, copying, and stealing information. Moreover, carelessness or communication errors can bring about errors or cause (unauthorized) changes to slip their way into an identity-file or profile. The consequences can be far-reaching: loss of creditworthiness, identity theft and being blacklisted for entry into the US.[48] Reports of identity theft make clear that a heavy burden of proof for the victims to prove that they are who they claim to be. Yet these risks associated with data technology have no obvious relation to the challenges of managing our moral identities in the context of IT, discussed in this paper. They are rather risks of *having* an identity in IT, making persons susceptible to cases of (unexpected or unforeseen) informational harm.[49] Notwithstanding the topicality and moral relevance of these issues, in this paper the focus is on three less clear-cut risks[50] associated with (constructing) identities in information technology.

2.6.1. Computational reductionism: the inevitable simplification

In the section on *super crunchers* I discussed the advance of a view expressing the quantitative or computational paradigm associated with such information

---

47 I have also referred to this as identification technologies. Supra note 26.

48 On several occasions, academic scholars were refused entry to the United States on the basis of their last names, see for example http://notthebeastmaster.typepad.com/weblog/2007/10/us-government-b.html, http://www.jihadwatch.org/archives/007169.php

49 This notion was coined by Jeroen van den Hoven (1998), referring to harm based on (available) information on a person.

50 Mireille Hildebrandt (2006, 2008) also warns against the unobtrusive effects of profiling on self-identity construction.

technologies. The exponential increase in the use and application of information technologies in all aspects of modern life presents us with access to ever-growing quantities of (raw) data that are at the very least challenging for our human capacities to make sense of them. The continuous accretion of information contributes to the need to have this information structured and processed, for example by means of super crunchers. This inevitably implies that despite the opportunities provided by exalted and sophisticated software, e.g. by making use of fuzzy logic and infinite decision trees, the rationale of this technology forces us to express information in terms of confined sets of data. It is static: thus, you have a 67 % chance of developing breast cancer type a. A convicted thief is (under such and such circumstances) 78% recidivist. By comparing your web surfing behavior to other customers, you are 22 times more likely to buy the book you are looking at today. The endorsement of the ideal that anything can be expressed in terms of data (and the probabilities and profiles based on them) ends in what I will call "computational reductionism."

Computational reductionism refers to the impossibility of processing machines to take into account *cognate*, soft information or data, such as contextual and motivational features, background knowledge, and (personal) explanation regarding actions or decisions. The resulting appearance of individuals as mere (statistical) objects does not match with the conception of individuals as managers of their own identities as individual and autonomous life-projects. This connects to the words of Isaiah Berlin:

> Rather than seek the liberty to do as I wish, what I may seek to avoid is [...] not being treated as an individual, having my uniqueness insufficiently recognized, being classed as [...] a statistical unit without identifiable, specifically human features and purposes of my own. (Berlin: 1969, 155)

In case of computational reductionism what is lacking is the first-person, self-informative perspective of the individual regarding her (moral) identity,[51] the perspective in which she sees herself in light of her ambitions, motivations, desires, life plan, and so on. Moreover, the element of indeterminacy (see Hildebrandt: 2006, 9) as part of the explorative construction of a self-informative identity is passed over.

[51] Jeroen Van den Hoven and I (2006, cf. chapter 4 of this thesis) argue for the incorporation of the first-person perspective into Identity Management technologies, through what we call "moral identification."

This implies a paradox for identities in Web 2.0, especially social networks: for one thing the users of these technologies are encouraged to present themselves and personalize their user domain, by implication an enhanced possibility of constructing and presenting oneself. For another the aim of these technologies and networks (besides social mapping) is to identify (and profile) these users, thereby taking them as objects of (algorithmic) computation.

2.6.2. Persistence of information

The second phenomenon in play where the discrepancy between explorative, self-informative identity construction and the imposed (re-)identification in IT finds expression is what I refer to as the "persistence of information." Information once captured in files and databases, and especially once it is out in the public domain e.g. through distribution on the Internet, information lasts, obdures: The ease of storing, retaining, copying, and retrieving data by means of information technology makes that information does not easily disappear.[52] It takes effort to remove. That is why information is "persistent." More so, it may rear its head at any unexpected moment. Think of a politician who, in the heat of running for president, is confronted with her leisure activities when she was a teenager. Relative to this, Jean-Francois Blanchette and Deborah Johnson (2002, 33–45) argue for the timely disposal of information as a fundamental requirement of informational privacy. This would preserve what they call "social forgetfulness, that is, allowing individuals a second chance, the opportunity of a fresh start in life."[53] For the explorative (see Mead: 1934) construction of self-informative identities individuals need to be able to wrest themselves from (former) characterizations and change in light of (new) moral considerations: "Privacy is the right to change, to reform your ways. The right not to solidify into what the register says you are," according to Ybo Buruma (2008).[54] Persistence of information confines the playground for "experiments in living" (see Mill: 1972, 271). By implication, this is an obstruction of the free and unrestrained, explorative construction of moral identities.

---

[52] With the exception of destroying (all) hardware, for example. Even so, on the Internet information is often so dispersed that it is not always straightforwardly gone. Anyhow, for the discussion I propose to set such scenarios aside.

[53] See also Mireille Hildebrandt (2008, 306), who speaks of losing the "right to oblivion" as a result of continuous dataveillance or data storage.

[54] Ybo Buruma cited in the report of the Rathenau Technology Festival 'Het Glazen Lichaam' (02-02-2008), available at www.scisa.eu/downloadfile.asp?ID=1362.

This phenomenon ties in with the identity management objection to data mining and profiling David Shoemaker puts forward elsewhere in this issue: "one has informational privacy when one has control over the access to and presentation of (unrevealed) information about one's self-identity" (see Shoemaker: 2010). A right to informational privacy implies, according to Shoemaker, the right to control or manage the presentation of one's (self-) identity, including its public construals. He claims: "I have a belief about how my self-identity ought to be presented." I agree, and would add that this is belief is not fixed; rather, it is subject to change. Therefore we should resist that former – and perhaps discarded – attributed identifications keep coming back to influence one's ID (management) process.

2.6.3. Dynamic nominalism

For the third phenomenon I draw on what Ian Hacking (1986, 236) calls "dynamic nominalism:" the co-emergence of kinds of human beings and acts, and our invention of ways to name them. It is more than just naming, it is "making up people" (1986, 236): "It is of more human interest than the arid and scholastic forms of nominalism, because it contends that our spheres of possibility, and hence our selves,[55] are to some extent made up by our *naming* [italics added] and what that entails." It starts from the idea that people implicitly shape their (moral) identities according to the possibilities and constraints offered by available labels, i.e. categories, profiles, attributed identifications. As a result, these labels or profiles are no longer just descriptive, but also constitutive. Hacking discusses how modes of description and possibilities for action relate to each other: "Hence if new modes of description come into being, new possibilities for action come into being in consequence" (see Hacking: 1986, 259). The increase in ubiquity and possibilities of profiling by means of information technology evidently enhances the aptness of dynamic nominalism; think for example of the coming into existence of new modes of identification as a result of belonging to a targeted consumer group. It is what Lawrence Lessig (1999, 154) calls "normalization:" "The system watches what you do; it fits you into a pattern; the pattern is then fed back to you in the form of options set by the pattern; the options reinforce the patterns; the cycle begins again." The phenomenon of dynamic nominalism also applies in worrisome cases like the seemingly proliferating pro-ana and pro-mias websites, respectively encouraging

[55] I would prefer to use the term 'identities' here.

anorexia nervosa and bulimia.[56] The rising availability of risk assessments in combination with probabilistic estimations and identifications of persons subsequently affect the ways in which persons see and come to see themselves. Consider the example of a person who is tested on the likelihood that she will develop breast cancer in her life. The test results show a .56 likelihood for her to become ill at age 35 and a .72 likelihood she may develop the disease at age 50. How does this information affect the construction of her (self-informative) identity? Whether or not she actually believes she will develop the cancer, she may start using this information for characterizing herself, e.g. as 'the woman who is (not) likely to have cancer in her life,' or the woman who is 'a potential cancer-patient,' or even 'the woman who once heard she had a reasonably large chance of developing breast cancer but who lived her life as if it wasn't so.'

Conversely, a profile may also cause a person to feel inhibited to try new things.[57] It may strike a 'young, ambitious potential' as inappropriate to enjoy leisure time at a fancy fair; the woman who thinks of herself as a potential cancer patient may feel reluctant to invest in a new home, or to make future travel plans.[58] Finally, the anorexia patient who is loyal to one of the pro-ana websites may identify so strongly with her pro-ana profile that it keeps her from getting treatment or realizing that she needs to do so. She is wholly engrossed or absorbed by her pro-ana profile; it constitutes her identity, i.e. *who she is*. Instead of autonomously constructing and reconstructing one's moral identity, the person fits herself to inferred categories.[59]

2.7. Conclusions towards...Value-conscious Identity Management

This paper discusses the advance of *super crunchers*, information technologies with the capacity of processing vast amounts of data. These technologies, I claim, further thinking in terms of a view in which everything seems determinate and tangible. At the same time, the development (and applications) of information technology presents us with a wide variety of means for constructing our moral identities. This results in a paradoxical situation, where on the one hand the

---

56 See for example work by Jenny Wilson *et al.*: 2006.

57 In connection with this, Sunstein (2001) points to the importance of unanticipated encounters.

58 Recently in the Netherlands, several healthy people were falsely diagnosed with Alzheimer. This caused them to live their lives in a way they would not have done otherwise, e.g. by selling their houses.

59 See also Daniel Solove (2004) on "normalization."

possibilities for 'identity management' in the moral sense of the expression seem enhanced by means of IT, yet on the other hand they seem increasingly constrained by the forensic logic that is inscribed in the technological structure of e.g. the 'Identity Management' technologies. In addition, the storage capacity and accessibility of information associated with the advance of IT limits our capability to get rid of certain attributed identifications in light of (new) moral considerations concerning our (moral) identities. The extent to which we are able to freely construct our moral identities is additionally constrained by the ubiquity of (probabilistic) profiles and categorizations by means of information technology.

The phenomena discussed in this paper demonstrate ways in which the structure of Identity Management (and other identification) technologies promote a presupposed forensic, nominal, practical notion of identity. This comes at the expense of a moral conception of identity which contains a reflexive, self-informative component of identification. On a conceptual level this paper hopefully contributes to a better understanding of the two senses of 'identity management' in the context of IT, and the different conceptions of identity that are involved. On a pragmatic level I suggest to take these considerations on board as values in the design of information technology, by means of "Values in Design." This term covers the collection of approaches concerned with integrating human and moral values into the design of our technologies,[60] known for example as "Value-Sensitive Design" (see Friedman: 1997, Friedman and Kahn: 2003), "Values at Play" (see Flanagan *et al.*: 2008), "Value-Conscious Design" (see Manders-Huits and Zimmer: 2009), and "Disclosive Computer Ethics" (see Brey: 2004). Its aim is to translate moral considerations into technical requirements for design. In case of 'Identity Management' as it is discussed in this paper, the design of the technology in question should provide for ways for the individual to construct and maintain their self-informative and nominal identities, in addition to their administrative, forensic counterpart. This includes for example integrating opportunities for experimenting with, reflecting about, and changing one's identity. Consider flexibility as a possible execution, for example by providing the user of the technology with the possibility for endorsing, adjusting, and shaping his or her profile. One could e.g. imagine the development of "identicons," i.e. icons similar to 'emoticons,' appearing on screen, indicating the source of the identity

[60] Lawrence Lessig (1999) has also pointed to the values in design and argued for the conscious design of our technologies. This also applies, I claim, to institutions.

related information used for this application, or the notification of other parties requesting to make use of this information.

To conclude, the expression of 'identity management' can be read in two different senses, which are related in the context of IT. The challenge from a moral point of view for designing identity related technologies with the epistemic aim to do justice to data subjects as moral persons, is to take into account the self-informative perspective that is part of 'identity management' as construed in the moral sense. In order to prevent (morally) problematic tensions concerning one's self-identity and -identification, designers of Identity Management technologies should be (made) aware of the moral understanding of persons as creators and managers of their identities and complement the forensic logic inscribed in information technology with flexibility.

2.8. References

Ayres, I. (2007) *Super Crunchers: Why Thinking-by-Numbers is the New Way to Be Smart*, Bantam Dell, New York.

Baumann, Z. (1993) *Postmodern Ethics*, Wiley-Blackwell.

Berlin, I. (1969) *Four Essays on Liberty*, Oxford University Press.

Blanchette, J-F. and Johnson, D. G. (2002) 'Data Retention and the Panoptic Society: The Social Benefits of Forgetfulness', in: *The Information Society: An International Journal* 18 (1): 33–45.

Bransen, J. (2008) 'Personal Identity Management', in: Mackenzie, C. and Atkins, K. (eds.), *Practical Identity and Narrative Agency*, Routledge, New York: 101-120.

----- (2000) 'Alternatives of Oneself. Recasting Some of Our Practical Problems', in: *Philosophy and Phenomenological Research* 60 (2): 381-400.

----- (1996) 'Identification and the Idea of an Alternative of Oneself', in: *European Journal of Philosophy* 4 (1): 1-16.

Brey, P. (2004) 'Disclosive Computer Ethics', in: Spinello, R. A. and Tavani, H.T. (eds.), *Readings in CyberEthics,* Jones and Bartlett, Sudbury MA: 55-66.

Borowski, E. J. (1976) 'Identity and Personal Identity', in: *Mind* 85 (340): 481-502.

Chaum, D. (1984) 'A New Paradigm for Individuals in the Information Age', in: *In IEEE Symposium on Security and Privacy:* 99-106.

Cocking, D and Matthews, S. (2001) 'Unreal Friends', in: *Ethics and Information Technology 2 (4):* 223–231.

Cocking, D. and Kennett, J. (1998) 'Friendship and the Self', in: *Ethics* 108: 502-527.

Copp, D. (2002) 'Social Unity and the Identity of Persons', in: *The Journal of Political Philosophy,* 10: 365–391.

Coven, V. (2003) A History of Statistics in the Social Sciences, *Gateway, Retrieved July, 15,* 2009, from grad.usask.ca/gateway/art_Coven_spr_03.pdf.

Davis, S. (2009) 'A Conceptual Analysis of Identity', in: Kerr, I., Steeves, V. and Lucock, C. (eds.), *Lessons from the Identity Trail*, Oxford University Press, New York: 205-218.

DeGrazia, D. (2005) *Human Identity and Bioethics,* Cambridge University Press, Cambridge.

De Vries, K. (2010) 'Identity, Profiling Algorithms and a World of Ambient Intelligence', in: *Ethics and Information Technology* 12 (1) 71-85.

Erikson, E. (1968) *Identity: Youth and Crisis,* Norton, New York.

Flanagan, M., Howe, D. and Nissenbaum, H. (2008) 'Embodying Values in Technology: Theory and Practice', in: Van den Hoven, J. and Weckert, J. (eds.), *Information Technology and Moral Philosophy,* Cambridge University Press, Cambridge.

Friedman, B. and Kahn, P. (2003) 'Human values, ethics, and design', in: Jacko, J and Sears, A. (eds.), *The Human-Computer Interaction Handbook,* Lawrence Erlbaum Associates, Mahwah NJ: 1177-1201.

Friedman, B. (ed.) (1997) *Human Values and the Design of Computer Technology,* Cambridge University Press, New York.

Gandy, O. (1993) *The Panoptic Sort: A Political Economy of Personal Information*, Westview Press, Colorado, Boulder.

Gates, H-L. (1992) 'Statistical Stigmata', in: Cornell, D., Rosenfeld, M., Carlson, D.G., (eds.), *Deconstruction and the Possibility of Justice,* Routledge, New York: 330-345.

Gergen, K. (1998) 'Erzahlung, Moralische Identitat und Historisches Bewusstsein. Eine sozialkonstruktionistische Darstellung', in: Jürgen Straub (ed.), *Identitat und Historisches Bewusstsein,* Suhrkamp, Frankfurt: 170-202. (english version "Narrative, Moral Identity and Historical Consciousness: a Social Constructionist Account" available at http://www.swarthmore.edu/x21118.xml).

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

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., and Krüger, L., (1989) *The Empire of Chance: How Probability Changed Science and Everyday Life*, Cambridge University Press.

Glymour, C. (2004) 'The Automation of Discovery', in: *Dædalus* 133: 69-77.

Glover, J. (1988) *I: The Philosophy and Psychology of Personal Identity*, Penguin Press, Goldman Harvey, London.

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

----- (1956) *The Presentation of Self in Everyday Life,* Doubleday, New York.

Hacking, I. (1987) 'Was There a Probabilistic Revolution 1800-1930?', in: Krüger, L., Daston, L., Heidelberger, M., (eds.), *The Probabilistic Revolution: Volume 1, Ideas in History*, MIT Press.

----- (1986) 'Making Up People', in: Heller, T. C., Sosna, M. N. and Wellbery, D. E. (eds.), *Reconstructing Individualism: Autonomy, Individuality, and the Self in Western Thought,* Stanford University Press, Stanford: 222-236.

Hand, D. J. (1998) 'Data Mining: Statistics and More?', in: *The American Statistician* 52: 112-118.

Harre, R. (1983) *Personal Being,* Blackwell, Oxford.

Heesterbeek, T. (2008) *U wordt verdacht.* Session report of the Rathenau Technology Festival "Het Glazen Lichaam" (02-02-2008), Retrieved, July 15, 2009, from www.scisa.eu/downloadfile.asp?ID=1362.

Heidegger, M. (1927) *Sein und Zeit*, Max Niemeyer, Tubingen.

Hert, de P. (2008) 'Beyond identity?', in: *Identity in the Information Society* 1: 123-133.

Hildebrandt, M. (2008) 'Profiling and the Identity of European Citizens', in: Hildebrandt, M. and Gutwirth, S. (eds.), *Profiling the European citizen. Cross disciplinary perspectives,* Springer, Dordrecht: 303–326.

----- (2006) 'Privacy and identity', in: Claes, E., Duff, A. and Gutwirth, S. (eds.), *Privacy and the Criminal Law*, Intersentia: 43–57.

Kamler, H. (1994) *Identification and Character – A Book on Psychological Development*, State University of New York Press, Albany, New York.

Kierkegaard, S. (1849) (alias Anti-Climacus). The Sickness unto Death. Copenhagen University Bookseller Reitzel's Press, in: Hong, H. V. and Hong, E.D. (eds.), Kierkegaard's Writings. Vol 6, Princeton University Press, Princeton, NJ, 1978.

Korsgaard, C. M. (2009) *Self-Constitution: Agency, Identity, and Integrity*, Oxford University Press, Oxford.

----- (1989) 'Personal Identity and the Unity of Agency: A Kantian Response to Parfit', in: *Philosophy and Public Affairs* 18 (2): 103-131.

Leibniz, W.G. (1969) *Discourse on Metaphysics,* Loemker.

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

MacIntyre, A. (1984) *After Virtue,* University of Notre Dame Press, Notre Dame.

Manders-Huits, N. and van den Hoven, J. (2006) 'Identiteitsmanagement en Morele Identificatie', in: *Algemeen Nederlands Tijdschrift voor Wijsbegeerte* 98 (2): 111-127.

Manders-Huits, N. and Zimmer, M. (2009) 'Values and Pragmatic Action: The Challenges of Introducing Ethical Intelligence in Technical Design Communities', in: *International Review of Information Ethics* 10: 37-44.

Mead, G. H. (1934) *'Mind, Self, and Society: From the Standpoint of a Social Behaviorist',* in: Morris, C. W. (ed.), University of Chicago Press, Chicago.

----- (1913) 'The Social Self', in: *Journal of Philosophy, Psychology and Scientific Methods* 10: 374-380.

Mill, J. S. (1972) On Liberty, Edited by H.B. Acton. Everyman Edition, London.

Nagel, T. (1998) 'Concealment and Exposure', in: *Philosophy & Public Affairs* 27 (1): 3-30.

Nissenbaum, H. (2001) 'How Computer Systems Embody Values', in: *IEEE Computer* 120: 118-119.

Parfit, D. (1984) *Reasons and Persons,* Oxford University Press, Oxford.

Perry, J. (forthcoming) 'On Knowing One's Self', in: Gallagher, S. (ed.), *Oxford Handbook of the Self,* Oxford University Press, Oxford.

Quetelet, A. (1831) Research on the Propensity for Crime at Different Ages. Translated and introduced by Sawyer F. Sylvester, Anderson, Cincinnati, 1984.

----- (1842) A Treatise on Man and the Development of his Faculties, translated by W. and R. Chambers, Edinburgh.

Raab, C. (2009) 'Identity: Difference and Categorization', in: Kerr, I. Lucock, C. and Steeves, V. (eds.), *Lessons from the Identity Trail: Anonymity, Privacy and Identity in a Networked Society,* Oxford University Press.

Ramdoelare Tewari, F. R. (2005) Identity Management Defined: How to position enterprises into the Identity Management framework, Master Thesis, Erasmus University Rotterdam, The Netherlands.

O'Reilly, T. (2009) What Is Web 2.0. O'Reilly Network, Retrieved 15 July, 2009, from http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.

Ricoeur, P. (1992) Oneself as Another, translated by Kathleen Blamey, University of Chicago Press, Chicago.

Schechtman, M. (1996) *The Constitution of Selves,* Cornell University Press, Ithaca.

Shoemaker, D. (2010) 'Self-Exposure and Exposure of the Self: Informational Privacy And The Presentation Of Identity', in: *Ethics and Information Technology* 12 (1): 3-15.

Solove, D. (2004) *The Digital Person. Technology and Privacy in the Information Age,* New York University Press, New York.

Strawson, G. (2004) 'Against Narrativity', in: *Ratio* (New Series) XVII (4): 428–452.

Sunstein, C. (2007) *Republic.com 2.0,* Princeton University Press, Princeton.

----- (2001) 'The Daily We. Is the Internet really a blessing for democracy?', in: *Boston Review,* Retrieved 15 July, 2009, from: www.bostonreview.net/BR26.3/sunstein.html.

Taylor, C. (1989) *Sources of the Self: The Making of the Modern Identity,* Harvard University Press, Cambridge MA.

Van den Hoven, J. (1998) 'Privacy and the Varieties of Informational Wrongdoing', in: *Australian Journal of Professional and Applied Ethics* 1 (1): 30-43.

Velleman, J. D. (2008) 'Bodies, Selves', in: *American Imago* 65 (3): 405-426.

----- (2006) *Self to Self,* Cambridge University Press, Cambridge.

----- (2005) 'The Self as Narrator', in: Christman, J. and Anderson, J. (eds.), *Autonomy and the Challenges to Liberalism: New Essays,* Cambridge University Press, Cambridge.

----- (2001) 'The Genesis of Shame', in: *Philosophy and Public Affairs* 30: 27-52.

Wilson, J., Peebles, R., Hardy, K., Litt, I. (2006) 'Surfing for Thinness: a pilot study of pro-eating disorder website usage in adolescents with eating disorders', in: *Pediatrics* 6 (118): 1635-1643.

3 Regulating Invisible Harms

This chapter is forthcoming as a book chapter in Simone van der Hof and Marga Groothuis (eds.), *The Impact of Europe on eGovernment*, Information Technology and Law Series, Asser International Press, 2010.

3.1. Introduction

As a result of the expansion of the public domain to a global level and the mobility within, there is an increasing need for identification and identity management. For this reason, Identity Management technology (hereafter IDM) is deployed. IDM is developed for the purpose of administration and management of user entities associated with information systems, networks and infrastructures. It forms the foundation on which access control is based (see Benantar: 2006, 2). A prerequisite for the use and application of IDM is the ascription of an 'identity' to each individual or actor for assigning duties and responsibilities. This identity is a unique referring token used for the identification of the information system user in question.

In the fairly standard account IDM is also sometimes referred to as "Access Management." In this sense it controls the access and restrictions of 'identities' (individuals as well as groups) to their rights, entitlements, opportunities, and accountabilities. In its most elaborate form, IDM encompasses the total of all means necessary in an organization or structure for electronic data processing, including user names, preferences, and access to services and applications. There is however a related, but broader sense in which IDM can be read, and that is the sense I am alluding to in this chapter. It also affects the registration and identification of persons; in this sense IDM is concerned with the collection of identity related data and the management hereof.

This chapter will take a closer look at the benefits and harms of deploying IDM for e-government. Though rarely recognized, IDM issues are connected to issues of justice: identification is closely related to the design and management of a liberal democratic society. Following from John Rawls' principles for the design of a just society, in order to be able to provide and safeguard freedom and equal opportunities for all, one of the tasks for government is to identify and

'manage' its citizens and their needs (see Rawls: 1971). The tension between personal interests and the common good (individual versus collective wellbeing) is precisely what is at stake in promoting justice and fairness in society; for example by knowing *who* is a potential threat to public safety or who is in need of special care or treatment. The (moral) responsibility of government to promote and maintain the common good of society, or what I call "care perspective," therefore requires the identification of (individual) persons or citizens.

The issue of free-riders in society, i.e. the temptation for individuals not to take their fair share in producing or arriving at a public good, warrants identification as well. Rawls explains how collective action and free-riding go hand in hand:

> [W]hatever one man does his action will not significantly affect the amount produced. He regards the collective action of others as already given [...]. If the public good is produced his enjoyment of it is not decreased by his not making a contribution. If it is not produced his action would not have changed the situation anyway. (Rawls: 1971, 267)

The apparent 'it doesn't matter what I do' in-effect of one's actions (in a large enough society) may lead people to evade their public duties or responsibilities. Rawls argues for governmental enforcement in this case: If we are all moved by the same sense of justice, the monitoring of such collective agreement should only be rational (see Rawls: 1971, 267-277). This by implication involves the identification of citizens; yet paradoxically it thereby also threatens certain public goods protecting the personal sphere of life such as privacy, trust, and freedom. I shall discuss this in more detail below.

The registration and regulation of identities and associated rights and duties have a long history; the need for identity management in the capacity of administrative systems in administration and government has been around for centuries. Well-known examples of identity management systems associated with government and administration are the national population's counts, such as the Census in the United States and the "Gemeentelijke Basisadminstratie" (GBA) in The Netherlands. In these systems, individual citizens are registered and uniquely identified by a number, a name, date of birth, and this information is replenished, depending on the system, by much more.

In the contemporary networked world, people interact with multiple identity-based organizations on a daily basis. There is also a growing trend in the

casualness and degree to which individuals are treated on the basis of their acquired identities; the treatment they receive, the things they are entitled to, their rights, accountabilities, the opportunities they are given and the limitations that are imposed upon them are shaped by the way their identities are construed and used.

IDM provides government with means for registering, monitoring, and communicating with citizens, in order to perform large-scale complex government service provision tasks efficiently and effectively. In doing so, there is a task for government to find a way of dealing with the tension mentioned above between the identification of citizens and their desired anonymity. Notwithstanding this challenge, IDM tools provide an effective way for communicating with citizens and executing governmental tasks.

This chapter is an appraisal of IDM for e-government. First I evaluate its success in terms of generating and structuring valuable knowledge for e-government practices in comparison with traditionally available epistemic tools. There are at least five good epistemic reasons for adopting IDM, as I will demonstrate by applying Alvin Goldman's criteria for evaluating the epistemic success of social practices. In what follows, I explore the possible risks for such 'informational' structuring of public administration and monitoring of citizens. I will focus in particular on the potential of what Joel Feinberg has coined "accumulative harm" (see Feinberg: 1984, 225-232), connecting to the virtually invisible risks of collecting and processing personal or identity related data.

3.2. The epistemic gains of Identity Management

Alvin Goldman has set out five criteria to evaluate how well different social practices lead to true beliefs. These criteria are power, fecundity, speed, efficiency and reliability. In this connection Paul Thagard (2001) has argued that the Internet is by its communicative nature a social practice invoking epistemic beliefs.[61] This, I would add, also holds for other IT-practices. The beliefs implicated in these practices can be about anything, and in the case of IDM they will mostly be beliefs about persons and their *identities*. Thagard points to the "veritistic" aim of Goldman's criteria, "presupposing that science aims at and sometimes achieves truth understood as correspondence between beliefs and the

[61] Jeroen van den Hoven also mentions Goldman and Thagard's work in connection with the Internet in 'The Internet and Varieties of Moral Wrongdoing' in: Langford, D. (ed.), *Internet Ethics*, MacMillan Press, 2000: 144.

external world" (see Thagard: 2001, 465-485). For the purpose of evaluating the epistemic success of IDM I would like to subscribe to this veritistic aim, holding the presupposition that what is sought in IDM are beliefs about a certain person (identity related data), corresponding with other information about this person in the external world, conceivably for other purposes. What we are examining here is the way IDM contributes to the acquisition of knowledge *about persons*.

3.2.1. Power

The first criterion for evaluation is 'power,' i.e. the ability of a practice to help people find true answers to the questions that interest them. In case of IDM and e-government, this interest comprises the registration of persons and the regulation of their rights and accountabilities. It finds expression in the execution of governmental tasks and responsibilities, that is by providing services to citizens in a broad sense; consider the communication of (e-)government with citizens and the providing for opportunities for their participation. IDM broadly construed possesses this epistemic power par excellence. The very nature of this technology is to enable the management of a legion of identities with associated rights and accountabilities and to make sure these identities (individuals as well as groups) find access to their entitlements.

3.2.2. Fecundity

The fecundity of a practice is its ability to lead to large numbers of true beliefs for *many* practitioners, i.e., many true believers. Needless to say, IDM improves the very fecundity in the sense of providing useful information about citizens and services to many government practitioners, since an IDM infrastructure enables multiple practitioners to access the same information at the same time. It is no longer necessary to demand information from other departments, causing delays in service and epistemic dependency on others to provide the correct and proper information.

3.2.3. Speed

The speed of a practice is how quickly it leads to true answers. Compared to the speed of getting the required information by using former methods of organizing and supporting governmental practices, the speed of using IDM exceeds these methods by far. Because IDM provides an informational infrastructure, enabling the linking of diverse kinds of information grouped in various preferred ways, the amount of time spent in organizing the required data

for each specific task is reduced to a minimum. Instead of setting up ways to getting the required information for each specific outset at the right time (and the organizational challenges that come with this), such time-consuming practices are eliminated and replaced by the use of an IDM infrastructure. What is more, the very speed of looking up information with the help of information technology exceeds the speed of humans doing the same.

3.2.4. Efficiency

Efficiency is how well a practice limits the cost of getting true answers. By reducing human effort in searching, organizing, and structuring information (that is, because of fecundity and speed), the direct costs of doing so are also reduced. However, the indirect costs such as the costs of designing, buying, and maintaining an IDM infrastructure also need to be taken into account. As a result, there is a break-even point above which IDM is far more efficient.[62] Especially for large institutions and organizations such as governmental institutions – for example at a European level – this point can be calculated and weighed against other mid- and long-term investments.[63]

3.2.5. Reliability

The reliability of a practice is measured by the ratio of truths to total number of beliefs fostered by the practice. In the case of IDM, this criterion is, similar to the case of the reliability of information on the Internet (see Thagard: 1997), perhaps the most challenging of all. In his assessment of the epistemic contributions of the Internet to scientific knowledge, Thagard shows that the Internet has severely diminished the mistakes associated with print or hand-copying. At the same time, the Internet has increased the possibilities for critical reflection on content and its associated revisions (see Thagard: 1997, 10). It has made information available on an even larger scale than the printing press - almost ubiquitous -, enabling continuous revision from a much wider public, as in the case of 'wiki's.' Yet this is also what constitutes a risk: The large-scale accessibility and the extended availability and spreading of information has increased the degree of 'data pollution.' For example when incorrect and poorly-formed data has been added to a database, this complicates its monitoring and

---

62 Below this point it may compete with other epistemic practices.

63 What we have not discussed here regarding efficiency is the term 'true' in its definition; however, this will be discussed below when we speak of the reliability of the acquired information.

review. Nonetheless Thagard emphasizes that knowledge-acquisition is a largely social enterprise (see Thagard: 1997, 10). This entails that although it has become more difficult to safeguard or monitor the quality of information, the increase in scale has also caused an increase in feedback from a multitude of perspectives, adding to its overall critical review.[64]

The epistemic reliability in case of IDM for e-government is also affected in another way. By communicating directly with citizens and enabling them to enter and/or verify their own personal information in government databases, the opportunities for unintentional mistakes vis-à-vis traditional government practices are diminished. Notwithstanding that citizens can also make mistakes (be it intentionally or not) when entering their own personal data, they are evidently better acquainted with this information and therefore (presumably) less likely to make mistakes. What is more, IDM provides for the possibility of linking this information to information already available on a data subject, for instance in other government databases, as a result of which the information can be verified.

In keeping with Goldman's standards of power, fecundity, speed, and efficiency, IDM surpasses former (non-technological) methods by far. The final criterion, reliability, also shows great advantages, yet calls for a cautionary attitude towards the quality of information in such informational infrastructures. In conclusion, the evaluation of epistemic success of information technology - especially IDM – for e-government demonstrates an overwhelmingly positive outcome in favor of deploying IDM. The remainder of this chapter will explore some of the harms or risks associated with the use of IDM for e-government. I will focus especially on what I call "accumulative informational harm," i.e. the harm resulting from the accumulation of seemingly harmless bits of information (-gathering).

3.3. Accumulative harm

Harms associated with any technology appear in many ways; they can be visible (overt) or invisible (covert), individual or collective, direct or indirect, diffuse, e.g. as the result of a chain of events eventually leading to harm. Moreover, harms

[64] This is especially relevant in open infrastructures such as the Internet, and perhaps to a lesser extent in secured infrastructures used for e-government; nonetheless, the entering, monitoring, review, and *linking* of information leads to similar issues concerning the quality of information.

are often not 100% likely to happen, in which case they are framed as probabilistic harms, or risks. In the case of the use of IDM for e-government, there are numerous harms we can imagine:

1. harm inflicted through malicious intent by government, e.g. the manipulation of an IDM infrastructure in order to exercise power in relations between government and citizens (for instance by excluding a certain group of citizens, or complicating access or procedures);
2. harm inflicted through recklessness by government, e.g. the deprivation of (a group of) citizens as the result of a careless implementation of IDM;
3. harm inflicted through malicious intent by the individual who enters the data, e.g. manipulation of the system by means of identity fraud;
4. harm inflicted through recklessness by the individual, e.g. registering incorrectly by mistake, thereby polluting the database and effecting future decisions regarding services.

These examples indicate that in most cases of harm there is an underlying assumption that the victims or perpetrators of harm are (ultimately) readily identifiable. Harm however, occurs regardless of us knowing who the victims or perpetrators are. In current discussions on intergenerational justice, (see Laslett and Fishkin: 1992; Page: 2006) harm that is possibly inflicted on non-existent, future (generations of) persons is one of the recurrent themes. If we assume that "actions or policies can only be wrong if they harm particular humans or non-human animals," now or in the future, this is referred to as an "identity-dependent" or "harm-based" account (see Page: 2006, 132-134). The problem for such an account is to deal with what is called the "non-identity problem,"[65] holding the view that there is no harm where there are no persons (directly) harmed. As explained by Edward Page (2006, 132) in the context of the intergenerational justice discussion, the actions or policies leading to harm are at the same time the "[...] necessary conditions of these people coming into

[65] For challenging cases on this topic e.g. resource depletion and the unborn child of a 14-year old cf., Derek Parfit (1987), *Reasons and Persons,* Clarendon Press, Oxford. See also Andrew Cohen (2009) 'Compensation for Historic Injustices: Completing the Boxill and Sher Argument', in: *Philosophy & Public Affairs* 37(1), 81–102 for a discussion of claims to compensation for a wrong that was also a condition of a person's existence.

existence."[66,67] It therefore involves, as put by Derek Parfit (1987, 351) "personal identit[ies] in different possible histories of the world."

In what follows, I discuss a covert, identity-*independent* type of harm: a harm not resulting from malicious intent or recklessness regarding a particular person or a collective, neither resulting from obviously wrongful or harmful conduct. What I will discuss as a possible harm in the context of IDM and e-government is what I call "accumulative informational harm." For this I draw on Joel Feinberg's notion of accumulative harm.[68] To introduce this notion, let me begin with a few examples:

> The first example is of a person walking on a nice, green city lawn.[69] The enjoyment of the person is harmless in that the grass may be (infinitely) slightly damaged but will recover quickly. However, if many people were to follow, the exact threshold depending on the season, the amount of rainfall, and so forth, the lawn would be damaged. The overall result might be devastating to all: There would be no lawn left.
>
> The second example considers the effect of exhaust gasses on our environment. The minor effect of exhaust emissions produced by one car may be considered negligible, yet the accumulation of exhaust fumes by multiple cars exceeds the threshold of harm (see Feinberg: 1984, 228) and causes substantial harm to our environment.
>
> Finally, Andrew Kernohan (1998, 73) portrays racism as an accumulative harm. Whereas one racist remark can be played out as marginal, the accumulation of racist remarks (and acts) is a serious harm: "[...]

66 To quote Derek Parfit: "It may help to think about this question: how many of us could truly claim, 'Even if railways and motor cars had not been invented, I would still have been born?'" (1987, 361).

67 To what extent this also applies to the harm I will discuss in this chapter will become clear in the discussion whether agents (victims and perpetrators) of harm are identifiable in this case, *see* p. 72.

68 Jeroen van den Hoven (2000) mentions Feinberg's notion of accumulative harm in "The Internet and Varieties of Moral Wrongdoing" in: Langford, D. (ed.), *Internet Ethics*, MacMillan Press, 153.

69 This example was borrowed from Andrew Kernohan, who uses it in his book *Liberalism, Equality, and Cultural Oppression* to point to the differences in act- and rule- utilitarian thought in attending to this problem. See Andrew Kernohan (1998), *Liberalism, Equality, and Cultural Oppression*, Cambridge University Press, 78.

> somewhere between those minor cultural acts and those produced by millions of people with racist attitudes the threshold of harm is reached."

Joel Feinberg (1984, 226) explains accumulative harm as a harm inflicted by a collective, through the accumulation of multiple seemingly harmless acts. Especially with regard to individual perpetrators, the (eventual) harm is seemingly invisible, non-existent. As Kernohan (1998, 72) writes: "[...] in the case of accumulative harms, a harmed condition can arise which does not result from harmful conduct." The concept of accumulative harm has proven useful with regard to several different social phenomena and harms, such as the use of antibiotics in agriculture (see Anomaly: 2009), environmental issues (water and air pollution) copyright infringement (see Moohr: 2003), and money laundering (see Alldridge: 2001, 279-319).[70] Andrew Kernohan (1998) provides a thorough review and application of the concept to the phenomenon of cultural oppression. Although cultural oppression is often neither noticed nor noticeable, by both victims and perpetrators, Kernohan successfully demonstrates why cultural oppression is nevertheless morally harmful and state regulation is warranted. His aim is to justify state intervention with respect to people's conceptions of the good on grounds of the resulting diffuse and insidious accumulative harms.

My purpose for this chapter is to show that there is a similar potential harm in play in the context of IDM and e-government. It is what I call *accumulative informational harm,* harm resulting from the availability and assembly of multiple seemingly innocuous bits of information about one person. Parallel to the notion of accumulative harm as described above, the harm is hardly noticeable - let alone noticed - by looking at the separate elements which together make up for the potential harm in question; as metaphorically put in the illustrious "cage" image of oppression by the feminist author Marilyn Frye (1983). The meaning of a cage is not grasped by studying its bars one by one, but by looking at the cage as a whole, i.e. at the *accumulation* of bars. The issue at stake thus concerns the potential harm created by the accumulation of multiple bits of sometimes seemingly innocent bits of (identity related) data.[71]

---

[70] Alldridge uses Feinberg's analysis of accumulative harm to demonstrate that criminal penalties are not warranted in case the contributions of a single perpetrator to the harmful effects of money laundering cannot be established.

[71] Interestingly, if we revisit what was said about free-riders on page 64, we find that the issue of free-riders, understood as an example of accumulative harm, does not only fuel the identification of citizens and therefore the deployment of technologies such as IDM;

This brings us to a difference with Feinberg's notion of accumulative harm: Whereas Feinberg's notion of accumulative harm refers to the accumulation of acts, acts that may *seem* harmless but turn out harmful in retrospect in their collective sense, accumulative informational harm concerns the potential harm *as a result of* accumulating bits of principally *harmless* information. It turns out that in the examples Feinberg provides e.g. of air pollution and the example Kernohan gives of racism, the individual acts were harmful all along, only in a negligible sense. In the case of accumulative informational harm, the particular bits of information are usually not considered harmful on their own, it is rather the accumulation of data with a potential for causing harm. This can be compared with issues of overpopulation: There is nothing harmful about a squirrel in New York; only above a certain threshold the clutter of squirrels proves to be a nuisance. And the same goes for the number of humans inhabiting this earth; there is principally nothing harmful about humans on this earth, however, overpopulation may lead to the depletion of natural resources (see Colby: 2009) and to a more rapid spreading of diseases (see Simonetta: 2009).[72]

Ethical theories, be it deontology, consequentialism, or utilitarianism, traditionally deal with issues concerning readily identifiable victims and violations, i.e. with agents and (harmful) acts. In the case of accumulative informational harm however, as we have seen above, none of these elements are readily present or easily identifiable; neither victims nor perpetrators are known at all times, as are the acts leading up to the harm in question.

Even so, something here is amiss in moral regard. Consider first that the ontology of traditional ethics is made up of identifiable agents. Conventionally ethical theory deals with the moral behavior of *known* actors: the identifiability of agents is a condition sine qua non.[73] For traditional ethics, the question is whether a *particular* agent's actions affect or have affected the well-being of *particular* ethical beings, for better or for worse. By implication these agents (both victims and perpetrators) are identifiable. As discussed, this is not necessarily true for accumulative informational harms. In case of accumulative

---

in fact, IDM brings on a certain type of accumulative harm itself. Thanks to Sabine Roeser for pointing this out to me.

[72] For an appraisal of world population control policies, cf., Matthew Connelly (2008) *Fatal Misconception: The Struggle to Control World Population*, Belknap, Cambridge.

[73] Another discipline where this can be clearly seen is game theory, which holds the basic assumption that players can be identified; there is no point in studying the behavior of unknown players.

harm the harm of a particular action may be negligible or (seemingly) non-existent. Although our purpose is to investigate a potential harm, neither perpetrators nor victims turn out to be identifiable. This conception of harm therefore falls under the category of (epistemologically) "identity-independent" harms, which Page has framed as repudiating the necessity for particular ethical beings as objects of harm in assessing whether it is wrong to perform certain acts or adopt certain policies (see Page: 2006, 138).

Moreover, traditional ethics deals with *acts*, with the implication of affecting the well-being of other (ethical) beings, for better or for worse. As pointed out by Walter Sinnott-Armstrong (2005), our moral intuitions appear to have evolved to primarily handle cases with obvious implications. And Edward Page (2006, 134) maintains that harm-based or identity-dependent reasoning is deeply ingrained in the ethics, law and commonsense morality of most countries. Hence the moral vacuum with respect to cases where people are unaware of the long-term or unforeseen, provisionally invisible effects of their acts. In case of accumulative (informational) harm, the harm is not even necessarily associated with its preceding acts. As Kernohan (1998, 72) puts it: "[...] in the case of accumulative harms, a harmed condition can arise which does not result from harmful conduct." It is indeed questionable whether the 'acts' of accumulative (informational) harm can be identified as such. Although the collecting of information or decision-making, for example to set up a database, could be seen as acts, what is often truly at stake is the information generated as a *by-product* of a certain act. Consider the case of a search act on the Internet by means of using Google; the by-product of this act is information *about* one's search behavior, which is recorded and stored in a database. On the basis of this information individual- and group-profiles can be made, in both cases a matter of producing identity related data, i.e. information about a person's identity. This newly produced information in turn adds to the collection of multiple bits of information about a person potentially causing accumulative harm.

What follows from this discussion is that the elements of traditional ethics, i.e. the acts and agents, are not straightforwardly meaningful with relation to accumulative informational harm. If neither acts nor identifiable agents are self-evidently involved, the question remains what makes this phenomenon count as harm. To answer this question let us take a closer look at the concept and objects of harm.

Harm[74] is generally defined as a disvalue, a detriment, or a set-back to socially valuable interests. These interests can be of two kinds, i.e. welfare interests, and interests related to one's personal projects and goals. Feinberg then identifies three ways in which someone's interests may be impaired: (1) the circumstances may be modified making it difficult to satisfy (competing) interests; (2) the degree to which prudential interests are protectively diversified is reduced; or (3) one's welfare interests are directly impaired making it difficult for someone to pursue the second kind of (ulterior) interests. For accumulative informational harm in the context of IDM and e-government this plays out for example as follows: (1) the opportunities for a citizen are modified on account of available information about him or her, e.g. profiles; (2) the range of options presented to the citizen to choose from (for different purposes) is limited as a result of available information or profiles; and (3) the pursuit of personal life choices is restrained as a result of the opportunities IDM provides for e-government.

Having addressed agents and acts as traditional elements of harm, and three possible ways in which harm can occur, what is left is a closer look at the objects of harm, i.e. *who* is affected. They are threefold: 1) individuals, 2) groups, and 3) culture.

Ad 1) The most straightforward object of harm, as discussed above in the context of traditional ethics, is the individual who the information in IDM is about. In this case, what is under consideration is the way the individual's particular interests may be set back as a consequence of certain acts (either by a group or by another person).

Ad 2) A group can also be the object of harm. For example in case of air pollution or cultural oppression, all members of the group are equally harmed. In this case the group is harmed on the basis of the shared characteristics of the group. Even if this group were eventually to consist of only one person, for example if only one person were left to be affected by the harmful consequences of air pollution or cultural oppression, the object of harm patently remains the

[74] According to Feinberg, we must distinguish between a non-normative notion of harm as a setback to interests, and a normative notion of harm as a wrong. Yet in *Harm to Others* (1984, 33) he offers a definition of harm as "a wrongful set-back to other people's interests" in which he conflates both conceptions, i.e. setbacks to others' interests that are wrongs at the same time. For a critical discussion on this topic cf., Heidi Hurd (1994) 'What In The World Is Wrong?', in: 5 *Journal of Contemporary Legal Issues* 157: 210-213, and Hamish Stewart (2001) 'Harms, Wrongs, and Set-Backs in Feinberg's Moral Limits of the Criminal Law', in: 5 *Buffalo Criminal Law Review* 13: 47–67.

group for the equal harm it has previously caused for other members of the group.

Ad 3) The third object of harm includes the creation of a potentially harmful environment. As an illustration let us take the accumulation of weapons in the United States:[75] The mere presence of weapons is not harmful as such. Yet it does create a strong potential for harm, in other words it brings about a hostile environment. Culture as the object of harm is paramount for any government-related context. It raises the issue of what kind of environment we wish to pursue. What is more, this object of harm is instrumental to the first two. By means of a harmed culture, both individuals and groups are or can be indirectly harmed.

Accumulative informational harm affects all three objects of harm. First of all an individual could be harmed e.g. as the result of incorrect identity related information stored in databases, or the improper treatment of a citizen on the basis of the incorrect application of a profile. Second, the technical hitches of a newly designed information infrastructure could (unintentionally) deprive certain groups or even all citizens of government services. Finally, the presence of elaborate digital files on citizens could contribute to a shift in power balance between citizens and government. The accessibility of personal information makes citizens more vulnerable, as is commonly known from (civil) wars such as in former Yugoslavia, World War II, and the Rwanda atrocities.

Priscilla Regan (1995) argues for privacy in connection with the latter object of harm concerning a change of environment. She argues that the problem with privacy is not that it harms individuals, nor members of a group by means of group characteristics; it is rather the "privacy-infringing" culture of facile data exchange precipitated by modern technology that harms the environment. Hence she characterizes privacy as a public good, valuable not only to the individual but to society in general. She points out that what is at issue here is how we collectively choose to organize society.

The contribution of the availability of elaborate (digital) files on citizens to a shift in power balance between citizens and government has proven of concern to many thinkers. One of the influential scholars in this field, Oscar Gandy, expresses discriminatory concerns as a result of the large-scale deployment of information technologies for the collecting, processing and sharing of data about individuals. In his book *The Panoptic Sort* (1993) he describes this mechanism as a "technology of power," exercising control over individuals through the sorting

75 Thanks to Jeroen van den Hoven for this useful analogy.

and self-sorting aspect of its functionalities. Gandy puts forward three worrisome developments with regard to the fast-pace maturing and associated integration of IDM and profiling technologies into everyday life: 1) the limitation and uneven distribution of available information, respectively options to choose from, 2) an increased instability in markets and politics due to the limited theoretical rationale of the systems deployed, 3) the destruction of trust and accountability within communities due to its totalitarian inclinations to include and conform individuals (see Gandy: 1993). According to Gandy, segmentation diminishes and eventually eradicates communication between different groups in society, thereby slowly undermining the public sphere and replacing it with multiple projected micro-experiences of a public life.

The discriminatory effects of profiling - and IDM technologies I would add - are also underscored by David Lyon, by what he calls "surveillance technologies." He warns for what can be seen as a new interpretation of the digital divide:

> To consider surveillance as social sorting is to focus on the social and economic categories and the computer codes by which personal data is organized with a view to influencing and managing people and populations. ... [I]n everyday life our life-chances are continually checked or enabled and our choices channeled using various means of surveillance. The so-called digital divide is not merely a matter of access to information. Information itself can be the means of creating divisions. (Lyon: 2003, 2)

3.4. Regulating invisible harms

Accumulative informational harm results from the accumulation of multiple bits and pieces of information. This information is made available for example by the data subjects themselves, or through the administration, tracking and tracing of behavior and characteristics of subjects, and by profiling. These informational processing techniques are closely connected with IDM and the opportunities such an informational infrastructure offers: the combining and linking of databases, the continuous updating of informational records, the structuring of massive amounts of available information by means of categorization, and so on. In short, the potential of accumulative informational harm comes with the very nature of implementing IDM (for any practice).

Having discussed accumulative informational harm as a risk of deploying IDM for e-government, what remains is how we deal with it. How do we

anticipate possible negative side-effects of the widespread collection, mining, and use of data in IDM technologies, whilst taking advantage of its epistemic (and organizational) benefits? Is this particular trade-off surmountable?

Value-Conscious Design (VCD) refers to a number of approaches purporting to meet such trade-offs in design (see Manders-Huits and Zimmer: 2009).[76] Would it be possible to avoid the phenomenon of accumulative informational harm by means of applying a VCD-approach, or is this particular kind of harm of a different order? Would it, for example, be possible to define thresholds for identity related information? Compare this with the working of a thermostat: it switches the energy supply off or on once the temperature reaches a certain critical limit. Would it be possible here to define such a limit?

After all, if we do not establish rules for dealing with our identity related data, our identities and individual biographies may be subject to the molding forces of macro-level institutional and cultural developments. Notably, this account is primarily an exploration of how we can responsibly take care of our social environment and the values within.

For this exploration I am sympathetic to the argument made by Priscilla Regan in *Legislating Privacy* that I have mentioned earlier, where she takes issue with privacy framed as an individual liberty (as opposed to a public good). She argues that privacy is instrumental to a democratic and just society, and the establishment of trusting relationships in such society. According to Regan, viewing privacy as an individual right, elicits a misplaced trade-off - amongst other things - between privacy versus public security. It then seems as if there is a trade-off between individuals giving up their privacy for the common good of public security, e.g. in case of tax evasion or crime-fighting. However, as a public good privacy is valued for its instrumental worth for democracy and thereby applies to all citizens alike as opposed to the comparison of particular inequalities associated with an individualist approach.

The framing of privacy as a public good rather than as a matter of individual rights, so I believe, corrects a misplaced trade-off between individual versus public values, such as privacy and public security in Regan's analysis. I think the

---

[76] For frameworks included under this heading cf., 'Design for Values', Jean Camp, (n.d.) 'Values at Play', Mary Flanagan, Daniel Howe, and Helen Nissenbaum (2008) 'Embodying Values in Technology: Theory and Practice', In Jeroen van den Hoven and John Weckert, (eds.), *Information Technology and Moral Philosophy,* Cambridge University Press and 'Value-Sensitive Design' Batya Friedman, Peter Kahn, and Alan Borning (2002) 'Value Sensitive Design: Theory and Methods', Technical Report, University of Washington.

key to resolving the trade-off between the potential benefits and harms of IDM for e-government as described in this chapter, lies in the prevention of accumulative informational harm, for example through the appreciation of privacy as instrumental to the design of our culture or social environment, while at the same time enjoying the epistemic and organizational gains of IDM for e-government.

Finally, what I propose as a starting point for the design and implementation of IDM technologies for e-government (and other purposes), and their regulation is a principle of minimalism: Record as little information as possible, for no longer as is strictly necessary, using a method or technology that is as robust as possible.[77] Though in practice this principle is in ongoing competition with economic and political forces such as control, security and risk aversion, I argue that it nonetheless contributes to avoid the following pitfalls:

1. The permanence of files and information they contain, irrespective, for example, of long-term developments in politics and policy. Information from records can be used and misused in the future for all sorts of reasons – think of the Second World War and how efficiently the Germans were able to do their detective work for demographic selection in the Netherlands, thanks to well-documented information on people's religious beliefs.
2. The issue of ownership of information within large infrastructures and organizations. In complex systems it is not clear precisely who owns what information, who grants access, who manages the file(s), and who performs the necessary checks.
3. The technical risks associated with large-scale information architectures, think of viruses, hacking, theft of information, carelessness.
4. The inescapability of a person's profile: There is no such thing as a 'clean slate.' Although someone may have changed over the years, policy decisions and treatments may still be based on a former profile.

[77] For many purposes, a combination of minimum information and minimum technology is sufficient. A successful example is the project *Verwijsindex Risicojongeren*. (VIR) It is part of a Dutch national information system meant to provide insight to different care providers regarding each other's involvement concerning a particular adolescent. For more information see http://www.verwijsindex.nl/ (available only in Dutch). What makes it a success is that there is no centralized and permanent database; the technology in question only supports the collaboration between associated parties for a clearly confined purpose.

5. The risks associated with incorrect information. The implementation of large-scale infrastructures reduces the risk of information not being available or being available in duplicate, but it increases the risk of the wrong information being available or information being incorrectly interpreted.
6. The opportunities for malicious intent by manipulation on the basis of available data,[78] e.g. identity fraud.

3.5. Conclusion

In this chapter I have discussed the overwhelmingly positive epistemic contribution of IDM for e-government. IDM adds to the epistemic power, speed, fecundity, efficiency, and reliability of e-government; moreover, it enhances organizational efficiency. On the other hand, the large-scale deployment of IDM for e-government also poses several risks, one of which involves accumulative informational harm. This potential harm is a result of the accumulation of multiple seemingly innocuous bits of information. Unlike traditional moral problems, neither victims nor perpetrators of this harm are readily identifiable. Nor are the preceding acts evidently harmful; the bits of information constitutive of the potential of accumulative informational harm are often by-products of other acts.

The challenge for thinking about values in design is to find out whether both the benefits of IDM for e-government can be kept, and harms prevented: Is this also possible in case of accumulative informational harm? One of the important considerations in this respect concerns the responsible design of our social environment. This is especially relevant in the context of e-government. An apposite starting point for the design of IDM for e-government is the principle of minimalism: It combines the ambitions and outspoken benefits of IDM whilst minimizing the amount of information needed in such design, so as to prevent reaching the threshold for accumulative informational harm.

[78] For an account of harm on the basis of information cf., Jeroen van den Hoven (2008) 'Information Technology, Privacy and the Protection of Personal Data' In Van den Hoven and Weckert, (eds.), *Information Technology and Moral Philosophy*, Cambridge University Press: 306-308.

3.6. References

Alldridge, P. (2001) 'The Moral Limits of the Crime of Money Laundering', in: *Buffalo Criminal Law Review* 289: 279-319.

Anomaly, J. (2009) 'Harm to Others: The Social Cost of Antibiotics in Agriculture', in: *Journal of Agricultural and Environmental Ethics* 22 (5): 423–435.

Benantar, M. (2006) *Access Control Systems: Security, Identity Management and Trust Models,* Springer, New York.

Budd, L. and Harris, L. (2009) *E-governance: Managing or Governing?*, Routledge, New York.

Cohen, A. (2009) 'Compensation for Historic Injustices: Completing the Boxill and Sher Argument', in: *Philosophy & Public Affairs* 37 (1): 81–102.

Colby, K. (2009) 'Current Human Population Dangerous for Planet', Retrieved October 15, 2009, from http://kevincolby.com/2009/04/01/current-human-population-dangerous-for-planet/.

Connelly, M. (2008) *Fatal Misconception: The Struggle to Control World Population*, Belknap, Cambridge.

Feinberg, J. (1984) *The moral limits of the criminal law, Volume 1: Harm to others,* Oxford University Press, Oxford.

Fishkin, J. (1992) 'Justice across the generations', in: Fishkin, J and Laslett, P. (Eds.), *The Limits of Intergenerational Justice,* Yale University Press, New Haven CT.

Flanagan, M., Howe, D. and Nissenbaum, H. (2008) 'Embodying Values in Technology: Theory and Practice', in: Van den Hoven, J and Weckert, J. (eds.), *Information Technology and Moral Philosophy*, Cambridge University Press: 322-353.

Friedman, B., Kahn, P. and Boring, A. (2002) Value Sensitive Design: Theory and Methods, Technical Report, University of Washington.

Frye, M. (1983) *The Politics of Reality: Essays in Feminist Theory,* Crossing, Trumansburg NY.

Gandy, O, Jr. (1993) *The Panoptic Sort: A Political Economy Of Personal Information,* Westview Press, Boulder Colorado.

Gaus, G. (1999) *Social Philosophy*, Sharpe Publishing, London.

Homburg, V. and Snellen, I. (2007) 'Will ICTs Finally Reinvent Government?', in: Pollitt, C., van Thiel, S. and Homburg, v. (eds.), *New Public Management in Europe: Adaptation and Alternatives,* Palgrave Macmillan, Basingstoke: 145-148.

Hurd, H. (1994) 'What In The World Is Wrong?', in: *Journal of Contemporary Legal Issues* (5) 157: 157-216.

Kernohan, A. (1998) *Liberalism, Equality, and Cultural Oppression,* Cambridge University Press.

Lyon, D. (Ed.) (2003) *Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination,* Routledge, London and New York.

Manders-Huits, N and Zimmer, M. (2009) 'Values and Pragmatic Action: The Challenges of Introducing Ethical Intelligence in Technical Design Communities', in: *International Review of Information Ethics* 10: 37-44.

Moohr Scott, G. (2003) 'The Crime of Copyright Infringement: An Inquiry Based on Morality, Harm, and Criminal Theory', in: *Boston University Law Review* (83) 731: 731-784.

Page, E. (2006) *Climate Change, Justice, and Future Generations,* Edward Elgar, Cheltenham.

Parfit, D. (1987) *Reasons and Persons,* Clarendon Press, Oxford.

Rawls, J. (1993) *Political Liberalism*, Columbia University Press, Columbia.

----- (1971) *A Theory of Justice,* Harvard University Press, Cambridge MA.

Regan, P. (1995) *Legislating Privacy: Technology, Social Values, and Public Policy,* University of North Carolina Press, Chapel Hill NC.

Snellen, I. (2005) 'E-Government: A Challenge for Public Management', in: Ferlie, E., Lynn, L.E. and Pollitt, L. (eds.), *Oxford Handbook of Public Management,* Oxford University Press, Oxford: 398-421.

Simonetta, J. (2009) 'Human Overpopulation: Causes, Effects and Solutions', in: *Ezine Articles*, Retrieved October 15, 2009, from http://www.ezinearticles.com/?Human-Overpopulation-Causes,-Effects-and-Solutions&id=1985200.

Sinnott-Armstrong, W. (2005) 'It's not my fault: Global Warming and Individual Obligations', in: Sinnott-Armstrong, W. and Howarth, R. (eds.), *Perspectives on Climate Change*, JAI Press.

Stewart, H. (2001) 'Harms, Wrongs, and Set-Backs in Feinberg's Moral Limits of the Criminal Law', in: *Buffalo Criminal Law Review* (5) 13: 47–67.

Thagard, P. (2001) 'Internet Epistemology: Contributions of New Information Technologies to Scientific Research', in: Crowley, K., Schunn, C. D. and Okada, T. (eds.), *Designing For Science: Implications From Professional, Instructional, and Everyday Science,* Erlbaum, Mawah NJ: 465-485.

Van den Hoven, J. (2008) 'Information Technology, Privacy and the Protection of Personal Data', in: Van den Hoven, J. and Weckert, J. (eds.),

Information Technology and Moral Philosophy, Cambridge University Press: 301-321.

----- (2000) 'The Internet and Varieties of Moral Wrongdoing', in: Langford, D. (Ed.), *Internet Ethics*, MacMillan Press: 127-157.

4 Identity Management and Moral Identification

This chapter is a joint publication with Jeroen van den Hoven, submitted to *Science, Technology, and Human Values*. It is a translation of 'Identiteits-management en Morele Identificatie', published in *Algemeen Nederlands Tijdschrift voor Wijsbegeerte*, 98 (2), 2006, 111-128.

4.1. Introduction

Globalization, mobility and international terrorism have been important factors contributing to the widespread desire for the unambiguous identification of individuals, and knowledge of 'who is who' on the Internet, concerning travelers, in shops and at the workplace. National governments and international trade and industry are investing heavily in methods to ascertain the precise identity of citizens and consumers and learn as much as possible about them. Both private and public sectors are using 'Identity Management' technologies for this purpose.[79] These technologies enable the 'managing' of identities in two ways. First, they regulate people's access to physical spaces, knowledge, information, communication infrastructures and a wide range of services by granting authorization on the basis of the established and verified identity of individuals. Second, Identity Management technologies enable the storage of elaborate digital representations of individuals in databases, to be processed and used for many different purposes. The way in which their identities are construed therefore co- determines the treatment that they receive, the things they are entitled to, their rights and responsibilities, the opportunities they are given and the restrictions or constraints that are imposed upon them.

Each individual in the Western world is represented in hundreds of databases, and these digital representations are assumed to be adequate for a rapidly growing number of applications. Categories, descriptions and models are used to represent people. More and more organizations act on the basis of the images of people that they obtain from files, data sets or profiles. Yet we know of the dramatic consequences that the availability of labels such as 'Jew,' 'Hutu,' 'Tutsi' (see Longman: 2001), 'white,' 'black' or 'colored' in administrative

[79] A Google search for the term 'identity management' yields roughly 32 million hits.

(identity management) systems have had for the persons to whom these labels were attached. Whether in political, military or racial conflicts – or even in routine practices in insurance and finance – profiles, postal codes, income data, educational records and behavioral indicators are used to sort and classify people (see Gandy: 1993, 2001). Government authorities understandably want to know who the tax evader is, or a serial rapist, or who could conceivably become one. They also want to know who is worst off in society and potentially in need of assistance. What are our obligations to persons when we represent, record and use their identities, or when we interact with them on the basis of these representations?

This article articulates the most important moral reasons for constraining Identity Management by persons other than those whose identities are at stake. First we discuss the type of identity that is implied in this context. We then discuss two fallacies in current thought on personal data and the protection thereof in light of the potential proceeding from modern Identity Management technologies. We propose extending the category of data related to persons and their identities that merits protection in two directions. By implication, we prefer the term "identity-related information" to "personal data." First and foremost, identity-related information need not necessarily be traceable to individuals in order to merit protection, as virtually all legislation concerning the protection of personal data presupposes; identification in a morally relevant sense can also take place without the referential use of descriptions. Nor do descriptions of persons or information about them need to be linked to the (possible) observation of individuals. Even without unique references and links to observations of individuals, Identity Management technologies can be used to develop practical and valuable strategies to find out more about people - whoever they may be - and to structure interaction with them in a digital environment. The structuring of interaction and communication is what is morally relevant.

In the second part of the article we put forward four moral reasons for protecting identity-related information in this broader sense. The fourth reason relates to moral autonomy, which calls for epistemic modesty with regard to the identifiability and claims of knowing persons, and respect for persons as 'self-presenters.' This leads to the formulation of the requirement of "moral identification," i.e. the incorporation of the first-person perspective regarding persons and their identities, by drawing on Bernard Williams' use of the term.

4.2. Biographical identity

Identity Management, as defined above, refers to 'identity' and 'identification' in a practical, biographical or forensic sense. John Perry (2002) refers to the biographical and forensic elements of our identities as 'objective representations' of persons. These are considered to be objective because their content is not dependent on who created them: they do not necessarily contain expressions using the personal pronoun 'I.' These representations are often created in part by people themselves. In that case they are the 'self-images' or 'self-representations' that people present of themselves. According to David Velleman (2006, 4), a person uses a self-image to represent "which person and what kind of person he is – his name, address and social security number, how he looks, what he thinks and feels, which traits define his personality, what principles he is committed to, and so on." Like Perry, Velleman views a representation of this kind not as intrinsically reflexive, since "it picks out the one he is, thus identifying him with one of the world's inhabitants" (see Velleman: 2006, 4). It is a way for us to imagine ourselves as potential referents of the pronoun 'who.' It is the representation of a person "considered non-first-personally but identified as the subject by some other, extrinsic means" (see Velleman: 2006, 7).

Identity Management relates not only to objective representations but also to unique references to individuals by means of descriptions, proper names, passwords, personal identification numbers and user names. We use these descriptions when we do not know or cannot see the people in question, i.e. when there are no *de re* thoughts. According to John Perry (2002) and Kent Bach (1987), thinking of a person by name or description can be reconstructed in terms of 'calling up a file on that individual.' They describe 'singular references' in terms of these 'files.' In Perry's view, we receive information about individuals in a role and we also act according to a role. Objects and people can play an 'agent-relative' role in people's lives, e.g. by 'standing in front of' or 'being on the phone with' them. These "agent-relative" roles are linked to epistemic and pragmatic methods, e.g. "methods for finding out about the object and methods for doing things to the object [...], the success of which depends on facts about the object" (see Perry: 2002, 224-225). Introducing yourself to someone and shaking his or her hand is an example of using information gleaned through the 'standing in front of' role, and this information also guides the action. This is all part and parcel of a standard information game that people play. Another important activity in this context is what Perry calls the "detach

and recognize" information game. He gives a business card as an example, but other examples of this game would be consulting a database at the town hall, a customer relationship database on a traveling salesman's laptop or a patient's medical record at a local hospital. When someone presents his or her business card, the objective representations (name, address, occupation, date of birth, CV, hobbies, etc.) are linked to the person who presents the card. When these parties go their way, the objective representations are separated from their observation. The next time we meet the person in question we can once again couple this information to an observation, identify the card holder and once again apply our beliefs to that person. This separation and identification procedure is crucial for communication where separate information is exchanged, so that others can link it to their own observations of the object (or person). The objective representations and their separation supplement are "agent-relative" perceptual ways of thinking and acting. Perry (2002, 226) stresses that "an objective representation with no possibility of being reattached to its source via some agent-relative role that supports pragmatic techniques is quite useless." It is often the case, for example, that beliefs about a person are stored in a number of mental 'files' because there have been different representations or self-presentations of that person (let us call them m1 and m2) with no belief that 'm1 = m2.' New beliefs about the person in the capacity of m1 therefore cannot be added to the collection of beliefs stored under m2 (see Bach: 1987, 43). As Perry (2002, 195) notes, "What unifies files, and makes two beliefs about the same person relevant to each other, is not that they are about the same person, but that they contain the same notion or linked notion."

The practical importance of unambiguous identification can be illustrated by the following attempt on the part of tax authorities to catch a tax evader (see Hill: 1997, 116-117). All taxpayers' bank accounts are registered under the account holders' 'real' names. Tax fraud, however, often entails having different accounts under *different* names. This way, the tax official does not notice the tax evader, as the names are non-substitutable. The ideal would be for two different names to be co-referential for one and the same account holder:

1. The tax authorities know that taxpayer X has more than $1,000,000 in the bank.
2. Taxpayer X has another $500,000 in an account under the false name 'Y.'
3. Since X = Y, the tax authorities know that Y has more than $1,000,000 in the bank.

(3) is false, of course, because 'X = Y' is not the same as 'knowing that X = Y.' Although it is true that the names 'X' and 'Y' are co-referential, they are not substitutable *salva veritate* in intentional contexts involving doxastic or epistemic attitudes such as believing or knowing. They refer in divergent ways to the taxpayer's identity, with the result that the tax fraud is not detected by the tax authorities.

This example illustrates the importance of having a coordinated way of recognizing and (re-)identifying people unambiguously. Understanding identification processes (e.g. the ways a person can be referred to) provides insight into a very important aspect of human existence. As Foucault pointed out some time ago, modern government cannot function without the ability to know who is who, keep people separate from one another, locate them physically and 'arrest' them if necessary, and link them to 'here-and-now files.' The two information games that Perry describes (introducing yourself and presenting your business card) and information on the substitution of co-referential descriptions are therefore crucially about 'managing identities.'

4.3. Data protection: restricting the use of identity-related information

Descriptive phrases can be seen as constitutive of partial identities in the sense that they can function as labels or content of mental, physical and electronic files on people: 'the man next door,' 'my closest colleague,' 'the man with the terrible aftershave,' 'John,' 'John Smith,' 'the man who always takes the 9:00 am train and alights at Central Station,' 'the owner of a blue Mercedes,' 'the person who paid 200 Euros using an electronic payment terminal in central Amsterdam on 1 August 2010 at 14:21:11,' 'the person on the surveillance videotape seen loading two orange boxes into the boot of a blue Mercedes,' 'the holder of account number 1234567,' 'the person in seat 55C on flight Q1 from Sydney to London on 2 October 2010.' These descriptions could all apply to different people, or they could all refer to the same person. If the latter is true they provide a great deal of information about John Smith, alias John.

'The owner of a blue Mercedes who lives at postcode 2345 XY' could refer to a number of individuals. In this regard, Keith Donellan (1966) distinguishes between the referential and attributive use of descriptions. The user of this description may have no specific person in mind: he may be thinking of the owner of a blue Mercedes, 'whoever that might be.' However, 'The owner of a

blue Mercedes' can also be used referentially if a specific person is being referred to. 'The man sipping his whisky' (said while pointing to a person at a party) is used in a referential way. The speaker's intention is to identify the person who – in his or her opinion – is drinking whisky, whether or not this is actually true. Even if the person in question is drinking iced tea rather than whisky and – strictly speaking – there is no one over there drinking whisky, the description still refers to the person drinking.

Descriptions used both attributively and referentially play a role in epistemic and doxastic strategies to collect information about people and expand our knowledge of them: both contain identity-related information. The police have a relatively large degree of freedom when it comes to entering information in files on suspects and collecting information about the identity of potential criminals: they can use biometric techniques such as fingerprinting and DNA evidence, data on Internet traffic, CCTV recordings, eye witness statements and any combinations of the above. The ethical and legal limitations involved in data protection and privacy laws are considered to apply primarily to descriptions of persons used referentially, but clearly both ways of using them play an important role in the surveillance of people and the management of their identities. International legislation on privacy and data protection defines personal data as follows:

> '[P]ersonal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity. (European Parliament: 1995, Article 2, Definitions)

The referential interpretation of 'identity' and 'identifiable' in this definition implies a narrow view of the moral limits to Identity Management, as descriptions used attributively may go unprotected. This seems to be a major oversight in data protection legislation, since we know that large amounts of data can be used attributively, e.g. in marketing and criminal investigation. We might, for instance, have a file on the owner of a blue Mercedes (whoever that might be) to which we add a long list of descriptions used attributively. The addition of just one small piece of information to this rich – but anonymous – data set could (much later) suddenly turn this data set uniquely referring, i.e.

'traceable,' to one individual. This is clearly illustrated by the anecdote Ruth Gavison puts forward:

> Consider the famous anecdote about the priest who was asked, at a party, whether he had heard any exceptional stories during confessionals. 'In fact,' the priest replied, 'my first confessor is a good example, since he confessed to murder.' A few minutes later, an elegant man joined the group, saw the priest, and greeted him warmly. When he was asked how he knew the priest, the man replied: 'Why, I had the honor of being his first confessor.' (Gavison: 1984, 352)

Did the priest breach the man's privacy in this case? Gavison claims that the first piece of information is anonymous and therefore does not in itself breach the first confessor's privacy. Although the second piece of information does apply to a specific individual, in the example it is communicated voluntarily by the person whose privacy is at stake and therefore cannot harm his privacy. Gavison (1984, 352) says that the second piece of information "turned what was previously an anonymous piece of information into further information about the individual." This is why she says there is a "translation from anonymous information to information about X."

Gavison presents this anecdote in the context of a series of problems associated with elucidating the notion of privacy. It all starts with the claim that "for a loss of privacy to occur, the information must be 'about' the individual" (see Gavison: 1984, 352). Given the prominence and importance of Identity Management techniques and technology, we ought to look at this claim afresh and, instead of defining the object to be protected in terms of descriptions used referentially, define it in terms of the broader concept of "identity-related information." This applies even if we are not talking 'about' a person referentially in the strict sense, because this information can be used in a digital environment for epistemic or doxastic strategies with the ultimate aim of finding out more about people.

Perry focuses on recognition and the importance of linking objective representations to observations in order to act ("recognition is a prelude to action," see Perry: 2002, 227). In his view, recognition entails comparing the characteristics of the person as observed with the characteristics already stored on that person. However comprehensive files and separate objective representations of persons may be, they must be linked to observations since all observations and actions are carried out by agents.

This is another pointer towards a restricted interpretation of the moral boundaries of identity management techniques such as recognition also take place at the level of 'files' of 'detached' representations, i.e. representations that are not linked to observations. Take, for example, an insurance agent responsible for handling and assessing theft claims. He knows the names and file numbers of the cases he is working on, and the one he is dealing with at present is a case of a stolen bicycle in a major city. During his lunch break he suddenly realizes that precisely the same person's bike was stolen last year too. The identities that the agent is processing are separate descriptions or objective representations, but in this case the appropriate type of recognition does not require 'reattachment' to an observation. Many cases of recognition and (re-)identification in information-rich environments relating to the identities of individuals are based on representations that are not – or not likely to be – linked to observations. The person in question is probably notified by mail or email, or sent an electronic bank statement and denied access to certain services accordingly. Many cases of identification in the context of Identity Management are based on descriptions used attributively that are not linked to observations in the way that Perry describes. They are attempts to identify anyone who fits the description, without it ever having been – or needing to be – linked to a particular person in an observational context in order to achieve the desired effect. In a digital world, identifications of this kind are an important element in epistemic or doxastic strategies to find out more about people or enable effective action to be taken. They cannot therefore be disregarded when considering the moral boundaries of Identity Management.

4.4. Moral reasons for data protection

We shall now examine four types of moral reasons for placing limits on the management of identity-related information or, to put it another way, four types of moral reason for protecting identity-related information.

4.4.1. Information-based harm

The first type of moral reason for considering moral restrictions on Identity Management by other people than oneself is to prevent harm that can be caused to individuals by making use of their identity-related information. Cyber-criminals and malicious hackers use computerized databases and the Internet to gain information on their victims and practice identity theft and identity fraud,

adopting the victim's identity and often causing financial loss and emotional harm in the process. In this way, people's bank accounts can be looted and their creditworthiness damaged, with the result that they are no longer able to obtain financial benefits and services. Stalkers and rapists use the Internet and on-line databases to track down their victims, which would often not have been possible if they did not have access to digital sources and were thus able to discover intimate details of their victims' lives. The information society has made people vulnerable to a new kind of harm that can be caused by identity theft or identity fraud, or simply by misusing identifying information. A principal justification for restricting the freedom of individuals causing, threatening to cause or likely to cause harm by using identity-related information is Mill's Harm Principle.[80] Protecting identity-related information reduces the likelihood of someone causing harm, similarly to the way that restricting access to firearms reduces the likelihood of people being shot dead in the street.

4.4.2. Informational inequality

The second type of moral restriction on the use of identity-related information relates to equality and fairness. More and more people are aware of the benefits offered by the market in identity-related information. When a consumer buys coffee in a modern supermarket, information on the transaction is generated, stored and added to their file or profile. Many consumers are beginning to realize that when they buy something at the checkout they have something to sell as well, i.e. information on the purchase or transaction, the "transaction data." Loyalty card programs are based on this principle. In the same way we can share information about ourselves on the Internet, where websites, browsers and autonomous software agents offer the prospect of paying us back in terms of increasingly useful information or discounts and services. Despite the fact that trade in personal data is flourishing, not all individual consumers are aware of this economic context. The rules of the game are therefore not clear. And even where this is the case it is not always possible to trade data transparently in a transparent and fair market environment so that people are paid an appropriate or reasonable price for it.[81] Moreover, consumers do not always know what the consequences of granting contractual consent for the use of identity-related information could be. We cannot simply assume that the developing market in

[80] "That the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others." (Mill: 1972, 16)

[81] This is particularly a problem on the Internet.

identity-related information will guarantee fair transactions (based on an independent criterion yet to be defined). Moral restrictions on Identity Management therefore need to be formulated so as to guarantee the equality and fair trade in identity-related information.

4.4.3. Informational injustice

A third moral reason – and a very important one – for placing restrictions on Identity Management by other people is based on Michael Walzer's concept of justice as complex equality. Walzer has put forward an argument challenging what he sees as the misleading simplicity of Rawls' interpretation of "primary goods" and his universal principles of "distributive justice." According to Walzer (1983, 8), "there is no set of basic goods across all moral and material worlds, or they would have to be so abstract that they would be of little use in thinking about particular distributions." Goods have no natural significance; their significance is the result of socio-cultural construction and interpretation. To ascertain what a fair distribution of a good is, we need to establish the significance of the good to those for whom it *is* a good. In the medical, political and commercial spheres there are different types of goods (medical treatment, political power and money respectively), which are allocated or distributed in different ways: medical treatment on the basis of need, political power on the basis of elections, and money on the basis of free trade or exchange. What needs to be avoided (and is in practice often prevented) is a situation where a particular good dominates. Walzer refers to a good as "dominant" when the individuals possessing it can influence the distribution of goods in other domains by virtue of the fact that they possess it (see Walzer: 1983). A monopoly is a way of controlling certain social goods and thus exploiting their dominance. In this case the benefits in one domain are translated one-on-one into benefits in other domains. A clear illustration of a situation of this kind is where money (the commercial domain) can be used to buy votes (in the political domain), preferential treatment in the medical domain or a university degree (in the educational domain) and so on. We resist the dominance of money and the influence that can be exerted with it when it comes to the distribution of other goods in other spheres. Moreover, says Walzer, we believe that political arrangements that permit or encourage such cross-contamination are unjust. No social good X should ever be allocated to people who possess another good Y merely because they possess Y, irrespective of the meaning of X (see Walzer: 1983).

What particularly upsets our sense of justice, according to Walzer, is (a) the allocation of goods from domain A based on the (normative) distribution logic of domain B; (b) transferring goods across the boundaries of separate domains; and (c) the dominance and tyranny of some goods over others. To avoid this, the "art of separation" of domains needs to be practiced and exchanges between them prevented: Walzer refers to this as "blocked exchanges." "Complex equality" obtains when the art of separation is practiced effectively, and the autonomy of the "spheres of justice" can be guaranteed (see Walzer: 1983).

Walzer's analysis can also be applied to information. The significance and value of information is local, and the distribution systems and local practices that provide access to it should therefore take local significances into consideration. The availability of information should be regarded as linked to particular spheres or domains. Many people have nothing against the use of their (personal) medical data for *medical* purposes, regardless of whether these relate directly to the personal state of health of themselves or their family, or possibly even with that of the community or the world at large, as long as they can be absolutely certain that the sole use being made of the data is *medical*, e.g. to cure disease. They do object however, when medical data is used to classify them socio-economically or to discriminate against them on the labor market, when they are refused commercial services (e.g. mortgages or insurance), or a political career is ruled out for them because of their medical history. People do not seem to mind so much if identity-related information on search behavior and interests is used by a library to improve *library services* to users, but they do mind if this information is used to ascertain – and possibly criticize – their character and tastes.

We therefore need to identify a third, important moral reason for restricting the use of identity-related information: informational injustice, i.e. a lack of respect for the boundaries of the "spheres of justice." What is often seen as an invasion of privacy would be more satisfactorily described as the morally inappropriate use of identity related data beyond the boundaries of what we intuitively regard as separate spheres of justice.

4.4.4. Moral autonomy and moral identification

Let us lastly consider the fourth type of moral reason for placing restrictions on Identity Management. This pertains to the moral autonomy[82] of persons, their right and ability to write their own moral biographies, reflect on their own moral development, make their own moral choices and evaluate them, without being exposed to the critical gaze of others and without feeling pressured to conform to 'normal' or socially acceptable identities. In this sense, a *morally autonomous* person is engaged in self-definition and self-improvement, and presents oneself as such to other people.

Modern individuals have cast off the idea of historical necessity and embraced the idea of contingency. In a changeable socio-economic environment there is a wide diversity of audiences to which individuals present themselves. According to the sociologist Goffman (1956), people keep these groups separated ("audience segregation") and present themselves differently to each one. Conflicts can arise, e.g. when a prospective son-in-law runs into his prospective mother-in-law on leaving his local pub with friends. In such cases self-presentations are put to the test and a certain amount of Identity Management is called for.

Establishing identity through other people's judgments and beliefs, and losing control over one's self-presentation sets up a barrier to experimenting with or in one's own life (what John Stuart Mill refers to as "experiments in living," 1972, 271). Modern individuals want to have the opportunity to decide their own morals, present themselves as they see fit and negate previous judgments and images based on new life experiences or fresh information. As Eugene Garver puts it:

> Contemporary freedom and choice go farther than Mill suspected – we all choose our identities, and make that choice from among a heterogeneous set of data, [...] we rarely choose our nationality, sex or religion, but we do choose to make these data part of our identity. (Garver: 1990, 391)

The idea of the person as a morally autonomous individual, the author and experimental subject of his or her own moral experiment, implies restricting other people in their attempts (directly or indirectly) to shape someone else's

[82] Joseph Kupfer (1987) makes a similar proposal. Privacy, he says, is a prerequisite for "[...] self-knowledge, self-criticism, and self-evaluation. This sort of control over self-concept and self is a second-order autonomy." (see Kupfer: 1987, 81-89).

identity, record it, publish it, or use it. Data protection laws should therefore provide protection against the recording of identity and identity-related information by anyone other than the person concerned. This is an argument in favor of requiring 'informed consent' for the processing of identity-related information.[83] How can we justify this principle of moral autonomy and its legal protection?

To explain why so much importance is given to respect for moral autonomy in the specific sense set out here, there are two aspects that need to be considered in our view: (1) the importance of respect for persons as self-presenters; (2) the importance of epistemic modesty when forming judgments and acquiring knowledge on the identity of persons.

(1) Velleman, in his analysis of privacy and shame, argues that an individual has a fundamental interest in "being recognized as a self-presenting creature, an interest that is more fundamental, in fact, than your interest in presenting any particular public image" (2001, 37). Events are shameful for individuals when they find themselves unable to present themselves as they would wish, thus undermining their status as self-presenting individuals (2001, 40). "When something private about you is showing, you have somehow failed to manage your public image, and so an inadequacy in your capacity for self-presentation is showing as well, potentially undermining your standing as a social agent" (2001, 38). Privacy norms, Velleman continues, are implicitly "norms of competence at self-presentation" (2001, 38). This also affects victims of stereotyping, who are characterized in a way that "leaves no room for self-presentation" (2001, 45). Someone who is characterized or stereotyped on the basis of ethnicity, for example, need not feel ashamed on account of his ethnic origin but rather because he feels represented as "less than the master of his self-definition and therefore less than a socially qualified agent" (2001, 45). Shame is "the anxious sense of being compromised in one's self-presentation in a way that threatens one's social recognition as a self-presenting person" (2001, 50). As Velleman's analysis of shame makes clear, interventions by other people in a person's Identity Management and self-presentation can compromise that person's status as a social actor.

[83] It goes without saying that there are domains where individuals cannot be allowed to write their own moral biographies from cover to cover. In this case they should at least be permitted to write those parts that are appropriate and be given the opportunity to authorize the parts that are, or need to be, written by others.

(2) Another explanation for the fact that we attach so much importance to moral autonomy lies in a requirement of epistemic modesty regarding claims to knowledge about persons. Actual knowledge of another person is always 'knowledge by description.' The person himself or herself, on the other hand, not only knows the facts of his or her own biography but is also acquainted with the associated thoughts, desires, emotions and aspirations. However detailed, comprehensive and accurate databases, files and profiles on a person may be, they are never enough to refer to a data subject in the way that the person himself or herself does, or could do. That person's knowledge and self-understanding can only be approximated in terms of objective representation. As Bernard Williams has argued, respecting a person implies the need to identify him or her in a special way. We use the term "moral identification" to refer to identification in Williams' sense:

> [I]n professional relations and the world of work, a man operates, and his activities come up for criticism, under a variety of professional or technical titles, such as 'miner' or 'agricultural laborer' or 'junior executive.' The technical or professional attitude is that which regards the man solely under that title, the human approach that which regards him as a man who has that title (among others), willingly, unwillingly, through lack of alternatives, with pride, etc. [...] each man is owed an effort at identification: that he should not be regarded as the surface to which a certain label can be applied, but one should try to see the world (including the label) from his point of view. (Williams: 1973, 236)

Moral identification presupposes knowledge – or at least recognition – of the data subject's point of view and an ability to empathize with what it is like for that person to live his or her life. People have aspirations, ambitions and expectations and see the things they think and do in that light. The representation of this aspect of people is precisely what is missing when identity-related information is stored in databases and they are dealt with on the basis of that information.[84] Identifications based on such data fall short when it comes to respect for the individual, in the sense that identities cannot be represented in the way they are experienced by the data subjects themselves. As Williams

[84] "[R]espect for human dignity relates to the need to avoid statistical dehumanization by undermining the identity of employees through data-processing techniques which allow for profiling of employees or the ranking of decisions based on automatic processing which concern them" (Council of Europe (1989), Explanatory Memorandum, 25, cited in Napier: 1992, 64).

observed in the passage quoted above, when we take an external, objectifying view of persons we are ignoring essential information about them. Understanding people 'externally' should ideally include not only objective representations of them but also what they want, wish or hope to be or to become, their gratitude, pride, shame and remorse and how they interpret this. The aim of the outsider's interpretation should be to represent and understand the person's second and first-order characteristics, which is undoubtedly difficult and may even be essentially impossible.

Emotions such as shame and attitudes such as shamelessness help to express what kind of person someone is: we take them into account when deciding what we think about that person. As Richard Moran points out (2001), it would be fallacious not to include emotions and attitudes in our considerations of a person, as this would be to misjudge the 'overall evidence' on him or her.[85] How a person reacts in terms of shame, pride, gratitude etc. constitutes fresh evidence on him or her that is morally relevant and forms part of the overall evidence on who this person is. To make do with less than this would be not only morally wrong but also epistemically irresponsible (see Moran: 2001, 182-183). Respect for persons, then, has an epistemic dimension as well as the aspect of respect for them as 'self-presenters.' Respect implies recognition of the fact that it is impossible to identify the other person with the person – and the representations – that he or she identifies with.

Even if we manage from time to time to get it right - by dint of exceptional empathy and attention - the question is still very much whether the data subject's experience of himself can be represented adequately in terms of the *dynamics* of moral persons. A person understands himself as someone who could improve morally, even if he makes no attempts in this direction, or only inadequate ones. A person cannot therefore be identified morally with something that is limited, bounded, fixed and unchangeable. This point is also made by the French existentialist Gabriel Marcel:[86]

> [I]l faudra dire que la personne ne saurait être assimilée en aucune manière a un objet dont nous pouvons dire qu'il est là, c'est-à-dire qu'il est donné, présent

[85] "[T]he apprehension of the mind of another person may thus only count as knowledge to the extent that it can approximate to this kind of awareness [...] [S]uch an approximation can never be more than a very distant one" (see Moran: 2001, 154).

[86] This is in line with the rule in French criminal law that statistical evidence relating to persons is not admissible in court.

> devant nous, qu'il fait partie d'une collection par essence dénombrable, ou encore qu'il est un élément statistique [...]. (Marcel: 1944, 31)

Isaiah Berlin (1969, 155) also suggests the importance of the subjective aspect of persons: "what I may seek to avoid is [to be] [...] insufficiently recognized, [...] a statistical unit without identifiable [...] purposes of my own." The person sees himself as 'work in progress' and on the move, as something that in principle could and should be improved, not a fixed reality but something 'under construction:'

> Elle se saisit bien moins comme être que comme volonté de depasser ce que tout ensemble elle est et elle n'est pas, un actualité dans laquelle elle se sent a vrai dire engagée ou impliqué, mais *qui ne la satisfait pas: qui n'est pas a la mesure de l'aspiration avec laquelle elle s'identifié.* (Marcel: 1944, 32)

As Marcel puts it, the individual's watchword is not *sum* (I am) but *sursum* (upwards). People have a tendency to be dissatisfied, giving them a constant aspiration to improve themselves. *Homo viator*, always on the move.

The parts of individuals' biographies that should be written, and the parts of their identities that should be 'managed' by other people call not only for the required identification and coordinated documentation and file management strategies that provide us with knowledge of who is who, they also demand attempts at *moral identification* as formulated by Bernard Williams: The person should be identified in such a way that he or she can (re-)identify with the constructed identity. This is not to say that an identity must necessarily be complete (for example that the tax authorities' database should contain additional information on a person's medical status), but it does imply an obligation to protect identity-related information and exercise restraint when using it, respect persons as 'self-presenters' and observe 'epistemic modesty' regarding persons and their identities.

In spite of these moral arguments for certain restrictions, information held in Identity Management systems is used extensively to profile people, i.e. to identify them using general profiles or develop such profiles. As soon as someone is being profiled in the first sense the data subject ought ideally to be involved, especially if the information is to be used at a later date – by somebody or other – for other purposes, as the following example illustrates.

A single woman in her early thirties is at the supermarket checkout. The woman in front of her realizes when paying that she has left her discount card behind, so she will not get the discount on the nappies she has on the belt. The first woman offers to help her and lends her own discount card, as she often does. Being classified in the system as a young mother,[87] she is subsequently sent advertisements for baby products by the supermarket. She is not particularly bothered; it is only when she is approached by various organizations in line with this 'young mother' profile (e.g. an agent for bank products offering her a family savings account or insurance) that it becomes an annoyance. The situation gets even more serious when this profile complicates her access to certain products or services. In this case harm or even discrimination is constituted, e.g. in case she is not presented with the options she would have otherwise received.

The point is that the person wants to retain control not only over the establishment of his or her own identity but also over that of identities constructed by other people. Respect for persons in the area of Identity Management, as argued above, calls for epistemic modesty and respect for self-presentation. Hence the pretension that individuals can be known completely, or at least adequately, which often seems to be implied in the practical implication of Identity Management technologies, understandably arouses opposition from the data subjects concerned, who resist this epistemic arrogation, often citing their 'privacy.'

Where we cannot leave it to individuals to write and edit their own biographies, simply because some facts about them need to be standardized (for example their dates of birth), they should have the right of authorization and correction whenever this is appropriate.

4.5. Conclusion

The availability of information technology is making people's identities the object of manipulation by people other than those whose identities are at stake. We have argued that all identity-related information – even descriptions that are used attributively rather than referentially, or descriptions that are not traceable to a unique, observable individual – merits protection, and that access to it should be made contingent on the informed consent of the person in question.

---

[87] This, of course, is a simplified and fictitious illustration of the effect and potential risks of a system of this kind.

We have articulated four moral reasons for protecting this information. Restricting the freedom of others to acquire or process identity-related information can be justified by considerations relating to: (1) preventing harm to the person or the data subject; (2) a fair and transparent market in identity-related information; (3) keeping the social spheres where identity-related information has different connotations and different practical significances separate; and (4) moral autonomy.

Respect for moral autonomy in our sense calls for (a) epistemic modesty in claims to knowledge about persons and (b) respect for persons as 'self-presenters.' The pretense that persons can be known completely or adequately rightly arouses opposition from the data subjects concerned, who resist this epistemic arrogation, often citing their 'privacy.'

Therefore, where we cannot leave it completely up to individuals to write their own autobiographies and design their own identities in Identity Management systems, since some facts about individuals need to be standardized and cannot be under the control of the subject, we argue that they have a right to authorize and correct, when and where appropriate. Moreover, the parts of individuals' identities that need to be managed by others for reasons alluded to above, require not only careful and thorough identification strategies for the sake of accuracy and completeness, but also require attempts at moral identification in Williams' sense. They should be encouraged to verify and manage their data, in order to keep control over their (perceived) identities and to prevent themselves of becoming 'numbers' in such Identity Management systems.

4.6. References

Bach, K. (1987) *Thought and Reference*, Oxford University Press, Oxford.

Berlin, I. (1969) *Four Essays on Liberty*, Oxford University Press, Oxford.

Council of Europe. (1989) Protection of Personal Data used for Employment Purposes, adopted by the Committee of Ministers of the Council of Europe.

Donellan, K. (1966) 'Reference and Definite Descriptions', in: *Philosophical Review* 75: 281-304.

European Parliament. (1995) 'Council Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free

movement of such data, adopted October 24, 1995', in: *Official Journal* 281: 31-50.

Gandy, O. (1993) *The Panoptic Sort*, Westview, Colorado, Boulder.

----- (2001) 'Dividing Practices: Segmentation and Targeting in the Emerging Public Sphere', in: Bennett, L. and Entman, R. (eds.), *Mediated Politics: Communication in the Future Democracy*, Oxford University Press, New York: 141-159.

Garver, E. (1990) 'Why Pluralism Now?', in: *The Monist* 73: 388-410.

Gavison, R. (1984) 'Privacy and the Limits of Law', in: Schoeman, F. (ed.), *Philosophical Dimensions of Privacy*, Cambridge University Press, Cambridge: 346-402.

Goffman, E. (1956) *The Presentation of Self in Everyday Life*, Doubleday, New York.

Hill, C.O. (1997) *Rethinking Identity and Metaphysics*, Yale University Press, New Haven.

Kupfer, J. (1987) 'Privacy, Autonomy and Self-Concept', in: *American Philosophical Quarterly* 24: 81-89.

Longman, T. (2001) 'Identity Cards, Ethnic Self-Perception, and Genocide in Rwanda', in: Caplan, J. and Torpey, J. (eds.), *Documenting Individual Identity*, Princeton University Press, Princeton NY: 345-357.

Marcel, G. (1944) *Homo Viator*, Editions Montaigne, Paris, Aubier.

Mill, J.S. (1972) *On Liberty*, edited by Acton, H., Everyman edition, London.

Moran, R. (2001) *Authority and Estrangement. An Essay on Self-Knowledge*, Princeton University Press, Princeton NJ.

Napier, B.W. (1992) 'The Future of Information Technology Law', in: *Cambridge Law Journal* 51: 46-65.

Perry, J. (2002) *Identity, Personal Identity and the Self*, Hackett, Indianapolis.

Schoeman, F. (ed.) (1984) *Philosophical Dimensions of Privacy*, Cambridge University Press, Cambridge.

Velleman, D. (2006) *Self to Self. Selected Essays*, Cambridge University Press, Cambridge.

----- (2001) 'The Genesis of Shame', in: *Philosophy and Public Affairs* 30: 27-52.

Walzer, M. (1983) *Spheres of Justice,* Basil Blackwell, Oxford.

Williams, B. (1973) *Problems of the Self,* Cambridge University Press, Cambridge.

5 What Values in Design? The Challenge of Incorporating Moral Values into Design

This chapter is forthcoming as an article in *Science and Engineering Ethics*, online first at http://www.springerlink.com/content/d28305n23v9715t3/.

5.1. Introduction: New technologies challenging ethics

Until the late twentieth century, technology was commonly regarded as value-neutral, i.e. mainly of instrumental value to human endeavours and activities (see, e.g., Florman: 1987). Recently, this conception of technology has been extensively challenged. One of the first prominent contemporary critics of the natural view of technology was Langdon Winner. He argued that technology is not value-neutral, but instead exhibits moral and political choices. He provided a compelling example of the highway overpasses built by Robert Moses, a famous and influential New York architect of urban planning (see Winner: 1980). In the 1930s Robert Moses was asked to design overpasses for the only highway connecting New York to Long Island, leading to Long Island Beach. Remarkably, Moses designed these bridges very low. The way they were built only allowed for cars to pass under whereas public transport (busses) passing under was rendered impossible. Supposedly, this implied Robert Moses favoured highways over public transport and community needs in his design. Some have argued that the highway overpasses were intentionally designed low so as to prevent public busses, the main means of transportation for the least well-off, including racial minorities, from getting to Long Island beach (see Winner: 1980).

Although there has been some dispute about whether this consequence was actually intended in the design (see Joerges: 1999), thanks to its strong illustrative power the example of Moses' low hanging bridges is often used to exemplify the possible political and moral import of design choices. Other examples of the inherent moral and political import of technology include the discussion of online search engines (see Zimmer: 2008; Introna and Nissenbaum: 2000), soft- and hardware codes (see Lessig: 1999), databases and classification systems (see Bowker and Star: 1999), cookies on personal

computers (see Elmer: 2004) and many more. A related value-laden phenomenon concerns the way in which information is presented, such as the presentation of risks associated with technology (see Asveld and Roeser: 2009) or the degree of reliability of information on the Internet (see Vedder and Wachbroit: 2003). The political and moral dimensions of technology have been recognized and have become important research topics in the philosophy and ethics of technology. Proponents of this account not only weigh and assess the risks and benefits of (new) technologies, but also address the issue of how technologies impact upon our moral decisions, actions, and ultimately, upon our lives. In light of the value-ladenness of technology it is highly desirable to have at our disposal means for ethically evaluating and justifying decision-making during technology design.

The increased awareness of possible implications of our design and our design choices for the expression, support, or undermining of our (political and) moral values contributes to an understandable desire to control and influence this process: Would it be possible to design our buildings, technologies, and institutions, so that they reflect, express, and enhance our moral and political views? Let us, for the purpose of this paper, set aside political values and focus on the practice of enhancing and incorporating the integration of moral values into the design of Information Technology (IT).[88]

In this paper I explore the criteria of adequacy for a methodology of ethics of technology while considering which normative considerations are implemented in design. I will consider and discuss Value-Sensitive Design (VSD), the most reviewed approach pertaining to values in technology design so far, as the prime candidate for such a methodology. First I describe the methodology as put forward by VSD in section 2. In the next section this approach is critically discussed from a conceptual, analytical perspective. The focus of this discussion is on the suitability of the VSD approach for integrating moral values into the design of technologies in a way that joins in with an analytical perspective on ethics of technology. From this follow the criteria of adequacy for an approach or methodology to implement moral values in design in the section 'Towards Value-Conscious Design.'

[88] The main methodological point of this paper however, holds for (values in) design in general.

5.2. Value-Sensitive Design

Human beings are confronted on a daily basis with the design, development and use, of (new) technologies that in some way impact upon them. Technological innovations in the field of IT have enabled the expansion of our social and political lives and activities to a global level - one of tremendous networks such as Instant Messaging services (MSN or Trillian), Skype, YouTube, MySpace, Flickr, Hyves, Twitter, and Facebook.

One of the prominent approaches for evaluating values in technology design – and information technologies especially - is Value-Sensitive Design.[89] VSD propounds a proactive approach regarding (the incorporation of) values in design, characterized by Van den Hoven (2005, 2008) as a way to "frontload ethics."

VSD emerged in the 1990s proceeding from the insights of "Human-Computer Interaction studies."[90] It started from the recognition that when designing information technologies, the predominant, traditional focus of engineers is on functionality; the primary interests of engineers concern usability, efficiency, reliability, and affordability of (new) technologies. An engineer's principal concern is to make a technology which has the required functionality. Yet there is more to technological innovation. As mentioned above, technology should not be considered to be value-neutral, but rather to have moral (and political) impacts on humans and their environment. Decisions made during the design process have value implications (see Van de Poel: 2009). The underlying idea is that technology is not merely enabling, but constitutive: it shapes our practices and institutions in important ways, such as those of health care and transportation, as Van den Hoven (2005) argues. Possibilities pertaining to (new) technologies are created as a direct or indirect consequence of design decisions whereas other possibilities are taken away. As a result, technologies can promote or undermine specific human values.

---

[89] Several authors in the field of values in design draw attention to human and moral values as an integral part of the conception, design, and development of technological artifacts and systems. These include Design for Values (see Camp: 2003), Values at Play (see Nissenbaum *et al.*: 2005; Flanagan *et al.*: 2008), Value-Sensitive Design (see Friedman: 2004; Friedman *et al.*: 2006), Values in Engineering Design (see Van de Poel: 2009) and Disclosive Computer Ethics (see Brey: 2001).

[90] For an academic overview of the field see Jacko, J. A. and Sears, A. (eds.) (2007), *Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (Second Edition), CRC Press, Boca Raton.

This is not to say values are solely designed into technology, nor solely conveyed by social drivers and forces. Influence is exerted in various ways, from the direction of the designers of a technology as well as from its users and stakeholders. When new technologies are designed and introduced, users may apply them for purposes other than those intended in design.[91] By doing so, technologies evolve, i.e. their functionality is adjusted and changed. This is what is referred to in VSD as an interactional position; design and social context and the interaction between both matter (see Friedman and Kahn: 2003).

Special VSD projects have been carried out for companies like Microsoft (on the reusing and sharing of information, such as code) and Intel. Numerous other projects have been taken up in the academic sphere, such as the RAPUNSEL project (involving the design of games for girls), projects dealing with informed consent online (via cookies and web browsers), network security, and projects in the field of human-robotic interaction or in the field of display technology. VSD assumes that:

> [h]uman values and ethical considerations no longer stand apart from the human-computer interaction (HCI) community [...] but are fundamentally part of our practice. This shift reflects, at least in part, the increasing impact and visibility that computer technologies have had on human lives. [...] (Friedman: 2004; Friedman and Kahn: 2003, 1178)

Technology may support/enhance and/or undermine/corrupt human values. For example, information available on the Internet increases access to and use of information, but might also entail the infringement of privacy or the dissemination of incorrect, false information. In order to do justice to these moral (and political) implications, VSD is employed as a methodology that "seeks to design technology that accounts for human values in a principled and comprehensive manner throughout the design process" (see Friedman *et al.*:

[91] Although the intended instrumental value of a technology or technological artifact is inherent to its intentional history (see Vermaas and Houkes: 2004), in complex technology development we are not able to foresee or predict the exact use and consequences of the (overall, final) technology due to uncertainty, i.e. a lack of information. This relates to what Anders Albrechtslund coins the "positivist problem," the fact that there is a problematic relation between intended design and final use. (see Albrechtslund: 2007) The use of a technology is not linear-causally determined by design, but plays out in practice, in a specific context of use. In consequence, there are many possible ways (multi-stability) in which a technology can be used in practice.

2000). What is more, VSD is a "way of doing ethics that aims at making moral values part of technological design, research, and development," says Van den Hoven (2005). In this paper this claim is investigated, by exploring whether VSD is indeed suitable as a methodology for designing values into technology. I start by describing the VSD methodology below.

5.2.1. Methodology

VSD is characterized by Friedman *et al.* as an approach that:

> [b]rings forward a unique constellation of features. First, Value-Sensitive Design seeks to be proactive: to influence the design of information technology early in and throughout the design phase. Second, Value-Sensitive Design enlarges the arena where values arise to include not only the work place [...] but also education, the home, commerce, online communities, and public life. Third, Value-Sensitive Design enlarges the scope of human values beyond those of cooperation [...] and participation and democracy [...] to include all values, especially those with moral import. Fourth, Value-Sensitive Design contributes a unique integrative methodology that involves conceptual, empirical, and technical investigations. [...] (Friedman *et al.*: 2002a, 2)

VSD evaluates and informs (the development of) technologies by taking into account human values. It sets out an integrative and iterative tripartite methodology, consisting of conceptual, empirical, and technical investigations (see, Friedman *et al.*: 2006 or Nissenbaum *et al.*: 2005). Each of the conceptual, empirical, and technical investigations and analyses are carried out iteratively, mutually informing and being informed by the other investigation. These interdependencies are metaphorically described by Nissenbaum *et al.* as "balls in the air:" Conscientious designers must juggle and keep in play the results of at least three modes, i.e. the results of empirical, conceptual, and technical research (see Nissenbaum *et al.*: 2005). The interaction among these three distinct analyses – and so the assembly of these separate methodological perspectives – is part of what makes the VSD an attractive enterprise.

The first methodological part involves a philosophically informed conceptual analysis. Its goal is to identify and articulate, on the one hand, the central values at stake in a particular design context, and, on the other hand, the stakeholders that are affected by this (technology) design.

Since this is its main goal or *raison d'être*, the lion's share of conceptual work in VSD evidently involves values. VSD is related to values in a two-fold way. First, it starts from the observation that design and technology may impact upon values: (new) technology developments may enhance, threaten, or transform existing values. For example, the advent of social networking sites has changed existing conceptions of privacy, especially for younger generations (see Livingstone: 2008). Secondly, values are fostered and built into design by means of VSD; it seeks to identify values considered to be of importance for a target group, say a certain company, society, or user group, and subsequently safeguarding these values by designing them into technology.

Friedman (2004; Friedman and Kahn: 2003; Friedman *et al.*: 2006) maintains that VSD particularly focuses on values with moral import that are often implicated in technological developments, such as the values of human dignity, justice, welfare, human rights, privacy, trust, informed consent, respect for intellectual property rights, universal usability, environmental sustainability, moral responsibility, accountability, honesty, and democracy. According to Mary Cummings:

> [w]hile neither independent nor exclusive, these values were selected in the development of VSD because they represent the broad values generally discussed in technology and ethics literature, as well as those that have become more important with the increasing use of computer technologies in everyday life. (Cummings: 2006)

She argues that VSD does not employ 'values' as referring to something being of economic worth, but that VSD takes up a broader sense of value: values refer to what persons, either singularly or collectively, consider as important to their lives.

To complement the value investigation, VSD sets out to identify the stakeholders of the technology in question. For VSD it is insufficient to solely articulate the central constructs of a design practice. This needs to be carried out recognising who is affected and to what extent. Direct and indirect stakeholders are distinguished; the former being those who interact directly with the new technology, the latter being those that are affected in a less straightforward way (see Friedman: 2004).

The first, rather abstract conceptual part of VSD's methodology is complemented with empirical investigation in the second part. Here the focus is on the way stakeholders assess the technology in question. Both qualitative and

quantitative research from the empirical sciences are used to inform this part of the deliberation process (see Friedman: 2004); focus groups, surveys, interviews, and measurements of user behaviour are some of the methods employed to investigate the stakeholders' assessment of the technology. The point of this research is to find out how stakeholders experience (new) technologies with regard to the values they consider important in relation to their social environment and reference groups.[92] Moreover, it concerns how they cope with emerging value conflicts. A primary consideration in this phase of analysis is investigating in what way design trade-offs affect perceptions, behaviours, and prioritization of competing values. This also includes the way in which the designer can support or detract from a value conflict (see Cummings: 2006). The results of these studies are applied for improving the design of (new) technologies by making them more sensitive and compliant with stakeholders' values.

Whereas the focus in both conceptual and empirical investigations is on the stakeholders' perspective with regard to who they are, the values they hold and the value implications of a particular (new) technology, the attention of the third part of the methodology – the technical analysis – is directed more specifically to the design and performance of the technology in question. As mentioned earlier, decisions during the design process knowingly or unknowingly determine to a large extent the way in which a given technology can be used in practice. Each particular design creates certain possibilities for action and application, while at the same time obstructing other possibilities. The technical part of the methodological investigation focuses primarily on how the technology can and will support, or compromise, the human and moral values identified in the other parts. This is where VSD has a pre-emptive stance with regard to ethics and technology: it purports to incorporate the results of the conceptual and empirical phases into design in a proactive way (see Friedman: 2004). As expounded by Van den Hoven (2005), "Value-Sensitive Design provides us with the opportunity to deal with [...] ethical issues in a new and fresh way: by 'frontloading ethics' and by all means the proactive integration of ethical reflection in the stage of design." The three parts[93] of the VSD methodology are

---

[92] "[V]alues cannot be motivated only by an empirical account of the external world, but depend substantively on the interests and desires of human beings within a cultural milieu," Friedman *et al.* (2006) say.

[93] These methodological parts can take part concurrently, though referred to by Flanagan *et al.* (2008) as "phases."

developed to secure the objective of not only making technology work, but making it work sensitive to human and moral values.

5.3. Will VSD do?

A discussion of VSD as a methodology for implementing values in design from an ethics of technology perspective

The potential of VSD as a proactive approach to ethics of technology is well appreciated. It recognises the importance of designing technology conscious of human and moral values over a mere retrospective perspective of discussing and dealing with value considerations after a technology has already been introduced and embedded in society. In contrast with retrospective analyses and the tailoring of technology to the requirements and needs of users with hindsight, VSD supports designing so that values are pre-emptively taken into account. What is more, VSD seems especially appropriate in dealing with a potentially diverse user population and the values they hold. However, in analysing whether VSD can meet the requirements associated with a (normative) methodology for implementing values in design (as part of ethics of technology), I argue that VSD falls short in a variety of ways that I will now discuss.

5.3.1. Who are the stakeholders?

Prior to the deployment of the VSD methodology, there is one basic issue that needs to be resolved first: How do we identify the stakeholders? This matter is of course not specific to VSD. In any stakeholder analysis the identification of stakeholders is paramount. (see e.g., Freeman: 1994; Mitchell, Agle and Wood: 1997) However, the objective of designing technology sensitive to - or even consciously and deliberately designed for – human and *moral* values, demands a much more exhaustive and comprehensive identification of stakeholders in order to do justice to underlying issues of fairness and equality. Whereas to meet the objective of designing a technology that is profitable, useful, or that warrants the inclusion of patently relevant stakeholders for a large enough group, incorporating moral values calls for further-reaching research in order to recognize who is affected by the technology in what way and to what extent.

This may seem relatively unproblematic at first sight, but in the case of more complex technologies, identifying stakeholders becomes increasingly difficult - let alone identifying the *indirect* stakeholders. This by implication raises doubts as to how the overall input provided by stakeholders, concerning their evaluation

of a particular technology, is to be valued or interpreted. What is the threshold for stakeholder input in order to obtain a sufficient and well-informed understanding of stakeholder values?

Once the issue of the identification of stakeholders is resolved, questions arise such as how these stakeholders can be reached, how their input is gathered and finally, how to deal with conflicting values. These are important issues and will be further addressed in the following sections.

5.3.2. How to integrate empirical methods?

The second point of attention concerns the use of empirical methods in VSD. VSD holds that it employs multiple methodologies of the social sciences to study how both direct and indirect stakeholders assess new technologies, i.e. the values these stakeholders render important in relation to a particular technology. The empirical investigation results in knowledge concerning the values and opinions of the stakeholders.

Consider the design of a medical system that supports decision-making for patients suffering from aneurysms. The system advises patients regarding their healthcare. In order to receive advice, the patients have to go online and fill out a questionnaire. What patients (and others) do not know is how the system comes to its conclusions. Therefore, the system is said to be 'opaque.' In addition, the patients are unaware of the available alternatives. Suppose we were to use VSD to evaluate and improve the values in the design of this medical system by means of empirical investigation. Now, imagine that one of the outcomes would be that a large part of the stakeholders fear that their autonomy is seriously threatened. What does this tell us and what can we learn from this information?

First of all, current technology development is often so complex that one can question whether stakeholders are in a position to fully assess new technologies. (New) Technologies are often so complex that they are unintelligible to most people, for example due to incomplete information availability or for lack of competence to assess the technology (how it functions, the possible effects on people, the consequences it yields for the environment, and so on). In the case of the medical decision support system described above, people may feel insecure because the system is opaque and they are unaware of possible alternatives. Moreover, people may err or have mistaken beliefs about factual and normative issues concerning the technology. Secondly, and more fundamentally, it is unclear what stakeholders actually intend to say when they are speaking of

particular values.[94] Not only is this dependent on the extent to which values are clearly defined in the (empirical) research, but also on the way the values are experienced and interpreted in each particular situation or assessment. Consider the value of autonomy in the example described above. Even if the definitions of values of stakeholders, designers, and researchers were to correspond, still these values could be interpreted differently, thereby generating different norms and actions.

Without giving ground and substance to values, interviewing stakeholders seems to be on loose grounds; as a result the values are too abstract and multi-interpretable, ultimately undermining the legitimacy of the empirical component of VSD. The outcome of research carried out by Friedman, Nissenbaum *et al.* (2002b) on users' conceptions of risks and harms on the web effectively illustrates this point. The study indicates that users emphasise security, privacy, and threats to computer systems as the key harms and risks associated with the web. These values and concepts, however, are not substantially defined in the research. Because, what does it mean if 67% of the users say that they fear security issues? And do they all refer to the same conception of security? In addition, how does the 67% fear of security problems relate to the 49% fear of privacy issues? Hence, despite the fact that the associated empirical research generates numerous results, Friedman, Nissenbaum *et al.* (2002b) acknowledge that it is difficult to draw conclusions based on them. Lacking in this research and in VSD in general, I claim, is reflection on how to deal with the (used) empirical investigations concerning values. As a result it is unclear how the integration of empirical studies relates to the overall – and especially the conceptual – investigation of VSD.

In addition, building on stakeholder opinions or evaluations as done in VSD implies a *shared* and *fixed* point of view:[95] it is assumed that stakeholders will *have* an opinion and that this opinion is relatively stable. However, people's opinions and beliefs (including moral beliefs) change because of new information, insights, and experiences. For example, people may be unaware of the implicit normativity of a technology until this is experienced, as a result of

---

94 This relates to the conceptualization of values, cf. the subsection 'What are Values?' (section 5.3.4)

95 This relates to the issue whether the context for deciding between options is 'static' or 'dynamic:' in static contexts, all options are known with a certain degree of probability regarding their consequences (see Schmidtz: 1995, discussed by Van de Poel: 2009). However, as pointed to before, uncertainty is inherent to complex technology (development) (see Sollie: 2007, 2009).

which they may come to disapprove. What is more, there is a plurality of values and normative positions in the public domain, and VSD needs to reflect on the issue of dealing with this, e.g. in case of persistent disagreement. Hence, this part of the methodology should be deliberative rather than taking the form of a survey so that issues can be made explicit and critically discussed.

5.3.3. The naturalistic fallacy

A third problem for VSD as a potential methodology for ethics of technology concerns the way in which choices for specific design options are (normatively) justified. As discussed, part of the VSD methodology consists of an empirical investigation, i.e. exposing the values of stakeholders involved. There is an implicit assumption in the methodology of VSD that one will know what to do *in a normative sense,* once these values are known. This, I contend, is where VSD runs the risk of committing the naturalistic fallacy, i.e. by reducing an 'is' to an 'ought.' Indeed, Friedman *et al.* explicitly state that

> [i]t is usually agreed (with Moore) that values should not be conflated with facts (the fact/value distinction) especially insofar as facts do not logically entail value. In other words, 'is' does not imply 'ought' (the naturalistic fallacy). In this way, values cannot be motivated only by an empirical account of the external world, but *depend substantively on the interests and desires of human beings within a cultural milieu*. [italics added] (Friedman *et al.*: 2006, 349)

Yet the conflation of facts and values is exactly what happens when the value stances of stakeholders are taken as the normative input for the VSD of a technology. The final (italicized) part of the quotation points exactly to the weak spot in this case: when Friedman *et al.* claim that values depend substantively on the interests and desires of human beings within a cultural milieu, this implies a sociological conception of values rather than an ethical one. Hence, this is a naturalistic interpretation of values that is running straight against Moore's account.

5.3.4. What are values?

Fourth, the concept of 'values' is central to VSD: VSD claims that human values and ethical issues do not stand apart from technology, but are fundamentally part of our technological practices. Technology, especially IT, has a rising impact on society and human lives; it may either support or undermine certain values.

As explained, the starting point for VSD is to identify stakeholders and values in technology (design), with the ultimate objective of incorporating values into technology by means of design decisions. Given that values are of paramount importance for VSD, I claim that more attention should be given to the conceptualisation of this notion.

As the starting point for exploring ethics and values in relation to technology design, Friedman and Kahn raise a series of questions. They ask,

> [I]f human values, especially those with ethical import [...] are important, they are no less controversial. What values count? Who decides? Are values relative? Or are some values universal, but expressed differently across culture and context? Does technology even have values? [...] [O]n what basis do some values override others in the design of, say, hardware, algorithms, databases, and interfaces? (Friedman and Kahn: 2003, 1178)

Although this outset seems promising from the perspective of ethics of technology, Friedman *et al.* remain descriptive with regard to the conceptualization of values. They describe a set of values "often impacted upon by technology." These are concepts that are certainly central to human existence, including major moral concepts such as autonomy, justice, and trust (see Friedman *et al.:* 2006). However, the definitions of these concepts remain abstract. Friedman *et al.* characterize the values they mention only briefly. Consider the value of autonomy, which Friedman and Kahn (2003) define as "people's ability to decide, plan, and act in ways that they believe will help them to achieve their goals." This (rather instrumental) definition of autonomy is formulated in such general terms that it can hardly function as a practical concept for implementing values in design from an ethics of technology perspective. Hence, Friedman *et al.* (2006) avoid the meta-ethical question about the nature of values just as much as the normative ethical question concerning which values matter, why, and to what degree. Maybe they do so because the ultimate conceptualization of each specific value should be left to each specific VSD.[96] Yet Friedman *et al.* remain rather vague on the conceptual level.

Illustrative of this is the way Friedman *et al.* (2006) discuss "UrbanSim," a project on integrated land use, transportation, and environmental simulation. They notice that stakeholders raise widely divergent values, and then ask how to prioritise and weigh these values. The solution offered in this case is a web-based

[96] I thank Ibo van de Poel for drawing my attention to this point.

interface in which stakeholders can "select indicators that speak to values that are important for them." (Friedman *et al.*: 2006) Notwithstanding the attractiveness and praiseworthiness of this undertaking, it seems VSD offers no methodological account for distinguishing genuine moral values from mere preferences, wishes and whims of those involved in the design process.

Notwithstanding this critique, Friedman and Kahn take a position with respect to values (in general) as they (2003) state that VSD defends what they call a "middle position" between moral universality and moral variability. Both extremes, they claim, are surrounded by problems: moral universality cannot find or ground its starting point for morality, i.e. a core set of values central to its moral conviction(s), whereas moral variability has problems regarding values shared among different groups or cultures (as a result of which values lose their meaning, cf. the problem of moral relativism). This leads Friedman and Kahn to adopt a middle ground, which:

> [a]llows for an analysis of universal moral values, as well as allowing for these values to play out differently in a particular culture at a particular point of time. [...] The general principle then is that designs need to be robust enough to substantiate the value under consideration and yet adaptable enough so that different cultures (or subcultures) can use the design in their own way. (Friedman and Kahn: 2003, 1183)

Flanagan *et al.* (2008) argue by the same token that:

> [t]he study of values in technologies pushes us to ask questions about origins and sources that are far from settled in the general study of moral and political values. In the absence of clear answers to these questions, recourse to a pragmatic middle-ground seems reasonable. This means going along with the idea of a hierarchy of values in which a 'thin' set are taken as common to all humanity and thicker sets associated with groups and subgroups of varying sizes and compositions [...] On the issue whether values can be derived, analytically, or must be discovered, empirically, a middle-ground sees virtue in both, drawing conceptual clarity and normative justification from theoretical works in moral and political philosophy, while supplementing these with knowledge about actual interpretations and value commitments of populations relevant to technologies under study. (Flanagan *et al.*: 2008, 326)

This position ties in with the method of Reflective Equilibrium, which has been put forward by Van den Hoven as most practicable for ethics of information technology (see Van den Hoven: 2008, referring to Norman Daniels: 1979; James Griffin: 1993; and John Rawls: 1971). Just like the middle position of VSD however, this method is contested due to its lack of clarity. It results in questions regarding the tenability of the 'thin' or universal values VSD puts forward: if values are universally valid for one thing but play out differently in various contexts for another, how can we avoid relativist arguments? Consider the value of human dignity: it is true that any culture can subscribe to this value. Yet what is left of its (universalist) meaning if the substance given to it differs in each context? In the one culture or context the value of human dignity may be taken as the basic principle to treat people equally, regardless of their race, skin colour, or religion. Yet in another context, the very same value may have an oppressing effect on people (even if unintentionally). Just consider historical and cultural examples regarding differences in dignity, e.g. of women vs. men, or 'black' vs. 'white' people. In the case of (new) technologies with global impact, such as the Internet, mobile communication, and airplanes, the differences among various interpretations of values become even more pressing. A technology that is introduced in two or more cultures with different conceptions of a value affects the moral emotions regarding the technology, including its uses and consequences, in different ways. For this reason I think that the middle position Friedman and Kahn adhere to, is subject to what is called "buck-passing:" at the end of the day, the problem is not solved but resurfaces elsewhere. What is lacking is the complementation of VSD with a justificatory theory of values, as will show below.

To recapitulate: the main criticism concerning the notion of values in VSD is that this notion remains underdeveloped. VSD should, in view of a normative approach for ethics of technology, explicate the difference between descriptive and normative values, i.e. what stakeholders factually consider important versus what they should regard as important. If this is lacking, the list of values runs the risk of becoming arbitrary, and any value serving the particular interests of designers (including the initiators and potential customers) may become a serious threat for VSD.

5.3.5. In need of ethical theory

The final point of attention for VSD from an ethics of technology perspective concerns the integration of ethical theory and analysis. Mary Cummings (2006)

suggests "VSD provides a road map [...] on how to include ethical considerations into design projects, even if they are not trained in philosophy or ethics." Although I agree that VSD *might* be able to include ethical considerations in design, I claim that in order to successfully meet this objective, VSD requires complimentary ethical theory and expertise.

Above, I address the conceptualisation of values. Now I will discuss the ordering and weighing of values: What values have priority over others? Are certain values paramount? From the literature it is not clear how VSD purports dealing with conflicting and incommensurable values (see e.g. Van de Poel: 2009). This leads to questions of how decisions should be made - and justified accordingly - in case of conflicting values. Notably, there are two sources of conflicting values: persistent disagreement between people (epistemological source) or dilemmas where trade-offs have to be made (ontological sources).

Now let us return to the example of the medical support-system; as mentioned this system is opaque, so we do not know exactly how it calculates or comes to its results. Nor do we, as patients, know the alternative treatments available to us, besides those suggested by the system. The design of such a (medical) system involves a variety of human values and associated questions, e.g.:

- Trust - is it possible to design a system that is technologically safe and sound, and that functions well so that it can be trusted both by physicians and patients?
- Privacy - is all the information collected necessary in order to generate the output, and is the database that contains the data about the questions and results protected against unauthorised access?
- Bias - is the medical system free from bias towards a specific preference for one treatment over the other; is the system not secretly a system produced or sponsored by a pharmaceutical company favouring one of their products?
- Autonomy - to what extent can the patient express his or her preferences regarding life and well-being?

In this example, as in most cases of technology (if not all), multiple values play a role. How do we compare and order these different values?

It becomes clear that VSD cannot be taken as a normative account in which competing values can be balanced and traded off as per a certain principle. In complex design situations, there is no clear-cut way to balance competing values

such as autonomy and trust, or freedom from bias and privacy. This raises an important question: who makes the final decision on how to prioritize these competing values? Is this left up to the majority of stakeholders to decide, i.e. when a certain value is ranked as important by a certain percentage of people, or is it left up to the designers? As argued by Van den Hoven (2010), we engage in ethics to come to reasoned solutions and clarifications to practical problems, such as value trade-offs. Although different conceptual ethical frameworks lead to different descriptions of the situation, ethical theory nonetheless provides for sources of moral arguments and moral considerations (see Van den Hoven: 2010). Therefore, I claim that VSD should be complemented with an explicit ethical theory.

5.3.6. Concluding remarks for VSD

VSD is an attractive approach for assessing values in (technology) design by means of an iterative process between conceptual, empirical, and technical investigations. In this section, VSD has been discussed as a potential methodology for ethics of technology. Despite its promising character, it turns out that VSD falls short in several respects. It cannot yet be considered as a fruitful methodology for implementing values in design from a normative perspective. Therefore, for VSD to be applied as an eligible methodology for ethics of technology, the issues raised in this article need to be dealt with first. At present VSD suffers from the following deficiencies: (1) VSD does not have a clear methodology for identifying stakeholders, (2) the integration of empirical methods with conceptual research within the methodology of VSD is obscure, (3) VSD runs the risk of committing the naturalistic fallacy when using empirical knowledge for implementing values in design, (4) the concept of values, as well as their realization, is left undetermined and (5) VSD lacks complimentary or explicit ethical theory for dealing with value trade-offs.

I claim that in order for VSD to function as a methodology for implementing values in design from an ethics of technology perspective, it needs to give more explicit attention to its value aims and assumptions. At this point the value analysis of VSD is descriptive rather than normative, leading to the question of whether Value-Sensitive Design can reach its original objective of enlarging the scope of human values to include all values, especially those with moral import. For the normative evaluation of a technology an explicit and justified ethical starting point or principle is required.

5.4. Towards Value-Conscious Design…

This paper was meant to explore the criteria of adequacy for a methodology of ethics of technology for implementing moral values into technology. VSD was taken as the prime candidate for such a methodology. However, what has been shown is that VSD lacks a normative, ethical component in order for it to meet the criteria for implementing (moral) values into design in an ethically justified manner. Taking values into account is not the same as normative reasoning. For this reason, I claim that applying VSD as it currently stands is a bridge too far for ethics of technology. For our purpose, technology should not only be made 'sensitive' to values, in other words, account for value considerations, but instead the objective should be to have technology consciously and deliberately designed to include ethical value considerations. I therefore propose to complement VSD with the following criteria:

1. The clarification and explanation of the overall (ethical) objective in view of a methodology of integrating moral values in design. In other words, the methodology to be used for implementing moral values into design needs to be explicit regarding its normative aims. What is the goal we wish to obtain? What entities are considered 'values,' is it desirable to make a technology sensitive to values in general, or primarily to moral values?
2. The next step is the explication of the ethical theory to be used. As discussed, VSD is in need of complementary ethical theory. By means of ethical theory, a light can be shed on ethical issues and associated moral arguments and considerations. To my view, making ethical deliberation explicit is crucial for a normative account; this way, relevant ethical concepts and value choices can be clearly identified and design decisions motivated accordingly.
3. Related to the previous points is the necessity for a clear-cut delineation of the (practical) objectives in each specific design situation with regard to the values to be considered in design. This includes a specification of *whose* values are to be considered (which stakeholders), what empirical methods are to be used, how, and why, and how does one come to points of action on the basis of this input without losing sight of ethical objectives (cf. (1)).
4. The safeguarding and monitoring of the process of incorporating moral values into design, e.g. by including a "values advocate" in the design team. (See Manders-Huits and Zimmer: 2007, 2009) This person is concerned with the effects of (1), (2), and (3) by (a) identifying and motivating choices for including certain stakeholders and leaving out others, (b) explicating value

conflicts and trade-offs, and (c) questioning and evaluating (design) choices in light of overall (moral) aims.

In conclusion, for an approach as VSD to become eligible as a normative account for integrating and implementing moral values into design, I propose to complement it with the explication of an ethical theory. This way, technology is not only made sensitive to moral values, but can also be consciously and deliberately designed that way.[97] I'd like to term this alternative reading of implementing moral values in design "Value-Conscious Design."[98] By means of engaging in Value-Conscious Design ethics is implemented in design, thereby activating ethical expertise for the benefit of improving our (technological) environment.

5.5. References

Albrechtslund, A. (2007) 'Ethics and Technology Design', in: *Ethics and Information Technology* 9 (1): 63-72.

Asveld, L. and Roeser, S. (eds.) (2009) *The Ethics of Technological Risk*, Earthscan, London.

Bowker, G. C. and Star, S. L. (1999) *Sorting Things Out: Classification and Its Consequences*, MIT Press.

Brey, P. (2001) 'Disclosive Computer Ethics', in: Spinello, R. A. and Tavani, H. T. (eds.), *Readings in Cyberethics*, Jones and Bartlett Publishers Inc: 51-62.

Camp, L. J. (2003) 'Design for Trust', in: Falcone, R. (ed.), *Trust, Reputation and Security: Theories and Practice*, Springer-Verlang, Berlin.

Cummings, M. L. (2006) 'Integrating Ethics in Design through the Value-Sensitive Design Approach', in: *Science and Engineering Ethics* 12: 701-715.

Elmer, G. (2004) *Profiling Machines: Mapping the Personal Information Economy*, MIT Press.

Flanagan, M., Howe, D. and Nissenbaum, H. (2008) 'Embodying Values in Technology: Theory and Practice', in: van den Hoven, J. and Weckert, J.

97 Notably the methodological comments of this paper are not directly related to technology; this methodology also applies to design situations apart from technology, e.g. institutions.

98 "Value-Conscious Design" was first introduced by Manders-Huits and Zimmer (2009) as an umbrella term to refer to the collective of initiatives involved with promoting human and moral values as an integral part of the conception, design, and development of technological artifacts and systems. In this paper the term is given more substance, subsuming the other initiatives and adding an ethical component.

(eds.), *Information Technology and Moral Philosophy*, Cambridge University Press: 322-353.

Florman, S.C. (1987) *The Civilized Engineer*, St. Martin's Press.

Freeman, R. (1994) 'The Politics Of Stakeholder Theory: Some Future Directions', in: *Business Ethics Quarterly* 4: 409-421.

Friedman, B. (2004) 'Value Sensitive Design', in: Bainbridge, W. S. (ed.), *Berkshire Encyclopedia of Human-Computer Interaction,* Berkshire Publishing Group: 769-774.

Friedman, B., and Kahn Jr., P.H. (2003) 'Human Values, Ethics, and Design', in: Jacko, J. A. and Sears A. (eds.), *The Human-Computer Interaction Handbook,* Lawrence Erlbaum Associates: 1177-1201.

----- (2000) 'New Directions: A Value-Sensitive Design Approach to Augmented Reality', in: *Conference Proceedings Of Dare 2000: Design of Augmented Reality Environments,* Association for Computing Machinery Press: 163-164.

Friedman, B., Kahn Jr., P. H., and Borning, A. (2006) 'Value Sensitive Design and information systems', in: Zhang, P. and Galletta, D. (eds.), *Human-Computer Interaction and Management Information Systems: Foundations,* Armonk, New York: 348-372. Reprinted (2008) in Himma, K.E. and Tavani, H.T. (eds.), *The Handbook of Information and Computer Ethics,* John Wiley & Sons, Inc., Hoboken, NJ: 69-101.

----- (2002a) Value sensitive design: Theory and methods (UW CSE Technical Report 02-12-01), University of Washington.

Friedman, B., Nissenbaum, H., Hurley, D., Howe, D. C. and Felten, E. (2002b) Users' Conceptions of Risk and Harms on the Web: A Comparative Study, *CHI 2002: Changing the World, Changing Ourselves,* Retrieved April 2, 2009, from http://doi.acm.org/10.1145/506443.506510.

Introna, L. D. and Nissenbaum, H. F. (2000) 'Shaping the Web: Why the Politics of Search Engines Matters', in: *The Information Society* 16 (3): 169-186.

Jacko, J. A. and Sears, A. (eds.) (2007) *Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (Second Edition), CRC Press, Boca Raton.

Joerges, B. (1999) 'Do Politics have Artefacts?', in: *Social Studies of Science* 29 (3): 411-431.

Lessig, L. (1999) *Code and Other Laws of Cyberspace*, Basic Books.

Livingstone, S. (2008) 'Taking Risky Opportunities in Youthful Content Creation: Teenagers' Use of Social Networking Sites for Intimacy, Privacy and Self-Expression', in: *New Media Society* 10 (3): 393-411.

Manders-Huits, N. and Zimmer, M. (2007) 'Values and pragmatic action: The challenges of engagement with technical design communities', in: Hinman, L., Brey, P., Floridi, L., Grodzinsky, F., and Introna, L. (eds.), *Proceedings of the Seventh International Conference of Computer Ethics: Philosophical Enquiry,* Center for Telematics and Information Technology, Enschede: 238-248.

----- (2009) 'Values and Pragmatic Action: The Challenges of Introducing Ethical Intelligence in Technical Design Communities', in: *International Review of Information Ethics* 10 (2): 37-45.

Mitchell, R., Agle, B., and Wood, D. (1997) 'Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts', in: *The Academy of Management Review* 22 (4): 853-886.

Nissenbaum, H., Flanagan, M., and Howe, D. (2005) 'Values at Play: Design Tradeoffs in Socially-Oriented Game Design', in: *Proceedings of CHI 2005:* 751-760.

Nissenbaum, H. (2005) 'Values in Technical Design', in: Mitcham, C. (ed.), *Encyclopedia of Science, Technology, and Ethics,* Macmillan, New York: 66-70.

Schmidtz, D. (1995) *Rational Choice and Moral Agency,* Princeton University Press.

Sollie, P. (2009) 'On Uncertainty in Ethics and Technology', in: Sollie, P. and Düwell M. (eds.) *Evaluating New Technologies: Methodological Problems for the Ethical Assessment of Technological Developments,* Springer, Dordrecht: 141-158.

----- (2007) 'Ethics, Technology Development and Uncertainty: an outline for any future ethics of technology', in: *Journal of Information, Communication and Ethics in Society* 5 (4): 293-306.

Van den Hoven, J. (2010) 'The Use of Normative Theories in Computer Ethics', in: Floridi, L. (ed.) *The Cambridge Handbook of Information and Computer Ethics,* Cambridge University Press, Cambridge.

----- (2008) 'Moral Methodology and Information Technology', in: Himma, K. E. and Tavani, H. T. (eds.), *The Handbook of Information and Computer Ethics,* John Wiley & Sons: 49-69.

----- (2005) Design For Values And Values For Design', in: *Journal of the Australian Computer Society* 7 (2): 4-7.

Van de Poel, I. (2009) 'Values In Engineering Design', in: Meijers, A., (ed.), *Handbook of the Philosophy of Science, Volume 9: Philosophy of Technology and Engineering Sciences,* Elsevier: 973-1006.

Vedder, A. and Wachbroit, R. (2003) 'Reliability of Information on the Internet: Some Distinctions', in: *Ethics and Information Technology* 5 (4): 211-215.

Vermaas, P. and Houkes, W. (2004) 'Action versus Functions. A plea for an alternative metaphysics of Artifacts', in: *Monist* 87: 52-71.

Winner, L. (1980) 'Do Artifacts have Politics?', in: *Daedalus*, 109 (1): 121-136.

Zimmer, M. (2008) 'Privacy on Planet Google: Using the Theory of "Contextual Integrity" to Expose the Privacy Threats of Google's Quest for the Perfect Search Engine', in: *Journal of Business & Technology Law* 3 (2): 109-126.

6 Values and Pragmatic Action: The Challenges of Engagement with Technical Communities in Support of Value-Conscious Design

This chapter is a joint publication with Michael Zimmer and is forthcoming in Emma Felton, Suzi Vaughan, Oksana Zelenko (eds.), *Ethics and Design*, Sense, 2010.

6.1. Introduction

While concern over the moral and ethical consequences of our modern technological era has existed for some time (see, for example, Friedman: 1997; Johnson and Nissenbaum: 1995; Moor: 1985; Nissenbaum: 2001; Shrader-Frechette and Westra: 1997), recent focus has been placed on how to develop pragmatic frameworks ensure that particular attention to moral and ethical values becomes an integral part of the conception, design, and development of emerging information systems (see Camp: n.d.; Nissenbaum *et al.*: 2005; Flanagan *et al.*: 2008; Friedman: 1999; Friedman *et al.*: 2002). Each of these frameworks – which we will refer to collectively as Values In Design – seek to broaden the criteria for judging the quality of technological systems to include the advancement of moral and human values, and to proactively influence the design of technologies to account for such values during the conception and design process.

We, the authors, have been involved in two separate attempts to engage with technical design communities to engage in Values In Design. Unfortunately, the results were discouraging. Drawing from our interventions, this chapter will identify three key challenges of pragmatic engagement with technical design communities: (1) confronting competing values; (2) identifying the role of the values advocate; and (3) the justification of a value framework. Addressing these challenges must become a priority if one is to be successful in pragmatically engaging with real-world design contexts to support the value-conscious design of emerging information technologies. To help illuminate how these challenges might play out in real-world design scenarios, we will contextualize them in

relation to a possible future application of Values In Design to address the privacy implications of Web search engine query data retention.

6.2. The move towards Values In Design

In *Technopoly*, Neil Postman (1992, 94) remarked how "we are surrounded by the wondrous effects of machines and are encouraged to ignore the ideas embedded in them. Which means we become blind to the ideological meaning of our technologies." It has been the goal of many humanists, social and philosophical scholars of technology to remove these blinders and critically explore the ideologies embedded in technical systems and artifacts (see, for example, Mumford: 1964; Winner: 1980; MacKenzie and Wajcman: 1985; Latour: 1992; Berg: 1998). These scholars argue that technologies have social and political biases; they tend to promote certain ideologies, while obscuring others. The concern that technologies have ideologies has been extended into questions of how information technologies specifically have *ethical* and *value* biases (see, for example, Moor: 1985; Johnson and Nissenbaum: 1995; Shrader-Frechette and Westra: 1997; Friedman: 1997; Nissenbaum: 2001; Tavani: 2004; Mitcham: 2005). Such scholarship sees information technologies as crucial sites for asserting not only social or political ideologies, but foremost human values of moral import, such as freedom from bias, trust, autonomy, privacy, and justice. This research seeks to identify, understand and address the value-laden concerns that arise from the rapid design and deployment of information technologies into society.

Arising from these concerns, various technical design communities have attempted to incorporate values into the design stages of technological systems. For example, the Human-Computer Interaction community has maintained a strong commitment to user-centered design toward the goal of increased usability of technical artifacts (see Norman: 1990; Nielsen: 1993; Raskin: 2000). Similar scholarship has emerged from Participatory Design communities (see Sclove: 1995), where democratic design processes are used to create not only efficient and effective technologies, but also ones that maintain the safety and well-being of users. While commendable for their longstanding commitment to incorporating values into the design of technologies, these frameworks tend to focus their efforts on functional and instrumental values, such as user-friendliness and worker-safety, falling short of directly addressing values of

moral import, such as privacy or autonomy.[99] To fill this void, new pragmatic frameworks have recently emerged to ensure that particular attention to moral values becomes an integral part of the conception, design, and development of technological artifacts and systems. These include Design for Values, Values at Play, and Value-Sensitive Design. Each of these frameworks – which we will refer to collectively as "Values In Design" – seek to broaden the criteria for judging the quality of technological systems to include the advancement of moral and human values, and to proactively influence the design of technologies to account for such values during the conception and design process.

6.3. Implementing Values In Design: Successes and disappointments

The collaborators who have developed various frameworks under the Values In Design umbrella have each enjoyed success in designing technologies with a focus on fostering particular values. For example, Friedman, Felten, and their colleagues developed web browser cookie management tools in support of the values of informed consent and user privacy. Similarly, Camp and her colleagues engaged in Values In Design (VID) by embedding the value of trust in web browser tools to protect Internet users from consumer fraud and identity theft. Flanagan and her colleagues are actively employing VID principles with the RAPUNSEL project, a computer game environment designed for teaching middle-school girls programming skills to help counter gender inequity in math and computer science while also embodying values such as cooperation, creativity, privacy and independence. And Howe and Nissenbaum's (2006) "TrackMeNot" web browser extension was designed to help obfuscate one's web search history records to prevent profiling by search engine providers, fostering the values of privacy and user autonomy. These examples reveal the promise of influencing the design of new information and communication technologies in order to account for moral and human values.

Encouraged by these successful implementations of VID frameworks, we, the authors, were optimistic as we attempted to engage pragmatically with the technical design communities of two emerging technological systems, hoping to influence their design in ethical and value-conscious ways.

---

[99] This is not to say that attention to functional values cannot have a moral or ethical impact. For example, building a user-friendly technology might increase a user's sense of autonomy. The distinction we are making here is a matter of focus, whether attention to moral and ethical values is the primary goal of the design process, or simply a by-product.

6.3.1. Vehicle Safety Communication technologies

Recent advances in wireless technologies have led to the development of intelligent, in-vehicle safety applications designed to share information about the actions of nearby vehicles, potential road hazards, and ultimately predict dangerous scenarios or imminent collisions. These vehicle safety communication (VSC) technologies rely on the creation of autonomous, self-organizing, wireless communication networks connecting vehicles with roadside infrastructure and with each other. In these networks, both vehicles and infrastructure collect local data from their immediate surroundings, process this information and exchange it with other networked vehicles to provide real-time safety information.[100]

Coupled with the predicted safety benefits of VSC applications, however, is a potential rise in the ability to surveil a driver engaging in her everyday activities on the roads and highways. VSC technologies potentially enable the collection of information on where drivers go, when they made their trips, and what routes they used. They represent a shift from drivers sharing only general and visually-observable information to the widespread and constant broadcasting of precise, digital information about their daily activities. The potential integration of VSC technologies into our daily activities on the public roads threatens to foster a scenario where we might be "driving into the panopticon" (see Reiman: 1995) of widespread surveillance, and the erosion of the values of privacy and autonomy as we drive along the roads (see Zimmer: 2005).

Since VSC technologies and their related technical protocols and communication standards were still in the developmental stage, Zimmer, with support from the PORTIA project,[101] attempted to apply a VID approach to influence the design of VSC technologies so that the value of privacy would become a constitutive part of the design process. Zimmer interacted with VSC project managers and engineers, was given access to technical white papers, and distributed his findings regarding the potential privacy and surveillance threats of VSC technologies. Unfortunately, the results of his engagement with the VSC design community were discouraging. Zimmer's efforts were met with skepticism: some designers viewed that the security gained through data

[100] For more information see U.S. Department of Transportation, and Vehicle Safety Communications Consortium.

[101] The PORTIA project is a five-year, multi-institutional, multi-disciplinary, multi-modal investigation funded by the National Science Foundation in the Information Technology Research (ITR) program. For more information, see http://crypto.stanford.edu/portia/.

encryption as a sufficient means of protecting user privacy, while others disregarded the threat to privacy altogether. Some potential privacy-protecting encryption solutions were rejected, reportedly due to concerns over efficiency and costs outweighing the privacy protections gained. Opportunities to engage further with the design community to address user privacy were hindered, as scheduled "privacy workshops" were cancelled by working groups, while changing personnel and legal concerns stymied sustained collaboration with more sympathetic designers.

While some increased awareness of the privacy concerns of VSC technologies emerged as a result of Zimmer's efforts, full and direct engagement with this design community to achieve the value-conscious design of these emerging technologies remained limited.

6.3.2. User Profiling Infrastructures

Online organizations are increasingly collecting data on users browsing and purchasing habits in order to create detailed user profiles, providing the tools to predict user behavior and provide personalized products and services. For example, online bookstores track and collect information about users' browsing and purchase history on their sites, resulting in the creation of detailed user profiles which allow the site to provide personalized purchase recommendations. Alongside this growing reliance on user profiles is the desire to be able to build, share and transfer profiles across various systems and platforms – the creation of a widespread User Profiling Infrastructure (UPI). For example, the user profile created on the online bookstore could also be shared with a different organization, like a movie rental company, to help predict what kind of movies the customer might want to watch. Similarly, a third party might be able to use a UPI to collect information on users' actions across various systems, such as a law enforcement agency monitoring purchasing habits across various websites in order to predict illegal activity.

While potentially useful, the cross-domain user profiling enabled by a UPI is fraught with value and ethical considerations, such as trust, informed consent, privacy and control over the flow of one's personal information. Recognizing the importance of addressing these value and ethical concerns, researchers from several Dutch universities and disciplines were invited to join the technical design team to help inform the development of such architecture. Manders-Huits was part of a team of academics specializing in ethics of information technology who participated in the project. Manders-Huits and her colleagues

participated in design meetings and discussions, did research in relevant ethical literature, and identified critical points for intervention to ensure that the UPI under consideration protect user privacy and other ethical concerns.[102] Overall, there was a concerted effort to engage in the value-conscious design of this technology.

This effort turned out to be more challenging than expected. While the technical designers were confronted with the challenges of the novel research field of ethics and technology, the ethicists found it equally challenging to apply ethical principles to the novelty and uncertainty of a still-emerging technical infrastructure. It was especially difficult to properly explicate and translate ethical considerations to workable requirements and specifications for the other project participants actually building the system. In other words, it proved difficult to operationalize the values so they could be put into the technical design.

More discouraging, however, was the insertion of a new project management team who failed to share the commitment to Values In Design or an appreciation for the demands such efforts bring to the design process. When it came time to speed up the progress of the overall project, eliminating (what were perceived as) the complex and time-consuming ethical concerns seemed the easiest route to get the project back on schedule. As a result, the value and ethical considerations of these profiling architectures were set aside.

6.3.3. Summary of implementations

While some of the original framers of the Values In Design (VID) frameworks have enjoyed success in designing technologies that sustain moral and ethical values, the attempts by Zimmer and Manders-Huits were disappointing and failed to achieve the desired results. The key difference between the successful implementations by the original framers of VID and the less-successful efforts by the authors is the site of engagement.

The successful applications of VID principles noted above share a common characteristic of being situated in the relatively non-hostile design environment of the academic laboratory. Each project's design team was, by and large, comprised entirely of academics committed to the goal of achieving Values In Design, creating a design context quite welcoming of each projects' goals. For example, one should expect little resistance against designing technologies that

[102] For more information, see https://doc.telin.nl/dscgi/ds.py/Get/File-52040/TUD_sotas.pdf

support human values when the design team is housed in Washington University's aptly named "Value Sensitive Design Research Lab." While the existence of such research centers is crucial for the continued development and refinement of VID methodologies, implementations that exist solely within their walls run the risk of being sheltered from the uncertainty of technical design projects and environments outside the walls of academia.

In contrast, Zimmer and Manders-Huits attempted to take the principles of Values In Design outside the laboratory and engage directly with designers of emerging technical systems in their native environments. Rather than benefiting from working within a sphere of designers already committed to designing for values, these imperfect engagements in real-world design contexts illuminate the challenges of moving VID from the academic lab to the industrial planning room.

6.4. Key challenges of Values In Design

Our unsatisfactory attempts to engage in Values In Design illuminate three key challenges that must be addressed before these pragmatic frameworks can be fully and successfully deployed in design contexts outside the academic laboratory: (1) the justification of a value framework, (2) identifying the role of a values advocate, and (3) confronting competing values. These three challenges are interrelated and not necessarily iterative, nor are they mutually exclusive, but we treat them as separate points for clarification and ease of discussion. To help illuminate how these challenges might play out in real-world design scenarios, we will contextualize them in relation to a possible future application of Values in Design principles to address the ethical implications of Web search engine query data retention.

6.4.1. Web search query data retention

Web search engines, like most websites, maintain detailed server logs tracking activity on their properties (see Google: 2005; IAC Search & Media: 2005; Yahoo!: 2006). Google (2005), for example, records the originating IP address, cookie ID, date and time, search terms, results clicked for each of the over 100 million search requests processed daily. Logging this array of data enhances a search engine's ability to reconstruct a particular user's search activities. For example, by cross-referencing the IP address each request sent to the server along with the particular page being requested and other server log data, it is

possible to find out which pages, and in which sequence, a particular IP address has visited. When asked, "Given a list of search terms, can Google produce a list of people who searched for that term, identified by IP address and/or Google cookie value?" and "Given an IP address or Google cookie value, can Google produce a list of the terms searched by the user of that IP address or cookie value?", Google responded in the affirmative to both questions, confirming its ability to track a particular user's (or, at least, a particular browser or IP address) activity through such logs (see Battelle: 2006a, 2006b).

The motivation behind retaining search query data is frequently justified in terms of improving search engine products and, by extension, user satisfaction. Google (2005), for example, states, "We use this information to improve the quality of our services and for other business purposes," while the search engine Ask.com explains, "We collect [...] anonymous information to improve the overall quality of the online experience, including product monitoring, product improvement, targeted advertising, and monetizing commercially oriented search keywords" (see IAC Search & Media: 2005). Such explanations describe the ongoing quest for the'perfect search engine,' one that would deliver personalized results based on a user's past searches and general browsing history, and deliver advertisements that are deemed useful or desirable for that particular user (see Hansell: 2005). Given a search for the phrase 'Paris Hilton,' for example, the previous query data retained by a search engine can help it determine whether to deliver results about the blonde heiress or a place to spend the night in the French capital, and whether to provide advertisements for celebrity gossip sites or Parisian bistros.

Despite these argued benefits, the practice of collecting and retaining search query data has not escaped controversy. Recent cases of the U.S. Department of Justice requesting search logs from the major search providers (see Hafner and Richtel: 2006; Mintz: 2006) and the release of user search histories by AOL without sufficient anonymization (see Hansell: 2006) have brought search query retention practices into a more public light, creating anxiety among many searchers about the presence of such systematic monitoring of their online information-seeking activities (see Barbaro and Zeller Jr: 2006; Hansell: 2006; McCullagh: 2006), and leading news organizations to investigate and report on the information search engines routinely collect from their users (see Glasner: 2005; Ackerman: 2006). In turn, various advocacy groups have criticized the extent to which Web search engines are able to track and collect search queries, often with little knowledge by the users themselves (see, for example, Electronic

Frontier Foundation: 2007; Privacy International: 2007), while both European and U.S. government regulators have started to investigate search engine query retention practices and policies (see Associated Press: 2007; Lohr: 2007).

In summary, while search engine providers claim users will benefit from the personalization of services retaining query data enables, privacy advocates worry about the potential harms of having users' online information-seeking activities routinely recorded and aggregated. Thus, we are faced with a kind of Faustian bargain: while personalized search services might be beneficial to users, they come with particular moral implications. Web search query retention, then, is an ideal site for pragmatic engagement using a VID approach. By engaging in VID, we can work towards the design of Web search engines that meet the needs of all stakeholders: provide personalized results for users who desire them, allow search engine providers to sell targeted advertising, while protecting user privacy. Outlining specific design principles to accomplish this is beyond the scope of this chapter, but by envisioning a hypothetical intervention with the search engine industry, we can provide a more concrete description of the pragmatic challenges that emerge in such interventions.

In the remainder of this chapter, we will discuss three pragmatic challenges for implementing moral values in design: (see Manders-Huits and Zimmer: 2009, Manders-Huits: forthcoming)(1) the justification of moral values, (2) identifying the need for a values advocate, and (3) confronting competing values. These descriptions will be augmented by a rendering of how each challenge might materialize in a hypothetical engagement with search engine designers to address the problem of search query data retention. Illuminating how these challenges might play out in a future real-world design scenario will help guide us to not only a richer understanding of each challenge, but also possible paths to their resolution.

6.4.2. Justification of value framework

It is important to identify whose value framework we are considering in each particular design setting. The challenge is to identify the extent or limits of the design community: Is it composed only of the actual technology designers, e.g. the engineers, or including contractors, managers, companies, potential customers, etc.? In other words, *whose* moral framework are we to study? After the identification of the design community and their values – as far as these are tangible – an important step can be taken with respect to value choices in the design process.

According to us, the aim of a particular VID project needs to be framed in terms of a desirable or preferred normative stance. This is how a technology becomes *conscious* of values. An organic, emergent, moral framework might be favored to justify the value choices to be encountered, or a particular (pre-)determined moral framework associated with a specific normative outlook. Either way, the starting point for the moral evaluation of a technology needs to be made explicit so that the value commitments of the technology are clear. By implication, the structure of the actual design process of this technology can be attuned to several points of intervention where value choices are made.

Friedman and her collaborators argue with respect to universal and particular values that a value-sensitive design should be flexible with respect to local values. The RAPUNSEL project provides a practical illustration: In this computer gaming design, players are tempted to take on part of the design of the game by choosing any particular set of preferences before entering into the game. A potential participant finds herself forced to choose between certain physical attributes, gender, race, etc. before entering the game. As she may not be comfortable with regard to any of these pre-designed categories, the designers allow significant flexibility to add nuances to the character. This flexible design feature, however, is easily provided from the safe settings of the academic gaming environment and we wonder, referring back, if this flexibility is equally manageable outside of these settings. For example, economic interests might constrain the number of options made available in a commercial gaming environment, and any such limitations present artificially simplify complexity of moral life.

In the case of search query data retention, the design of typical search engine interfaces prioritize simplicity and usability, featuring only a text box for entering keywords, and a search button to initiate the search. Users are for example not presented with options to determine if they want the search logged by the provider in order to help personalize future results or advertising. Perhaps some of them would choose to forego those benefits to protect their privacy, while others would not. However, current search engine design does not allow for this decision at the user level, predetermining the ability (or inability) for users to make their own value-driven choices.

The justification of value choices during design demands elucidation of the moral framework being used with respect to values such as freedom from bias, autonomy, privacy, and equity. Moreover, it calls for a different way of doing ethics. Where traditionally, ethics is a field where past events are discussed and

evaluated in order to develop a comprehensive account of all moral phenomena, what is required for a successful implementation of moral theory and values into design, is the "frontloading of ethics" (see Van den Hoven: 2005), also referred to as the changing of a retroactive to a proactive application of ethics. This is a challenging enterprise. Since technology design cases are fundamentally uncertain and with complex future outcomes, providing an operable account for ethically evaluating and guiding value choices is difficult. It calls for the (re-)consideration of traditional ethical accounts in light of the complexity of technological design. This however, goes beyond the scope of this chapter. More practically, a first step in the "frontloading of ethics" is establishing the role of a values advocate on technical design teams. However, as the next section will detail, this is often a complicated and contentious task.

6.4.3. Role of the values advocate

The various methodologies for implementing moral values in design suggest that design teams are to be complemented with scientists from other disciplines than engineering, from social sciences or humanities, for example, with the purpose of attending to the value commitments of the designers as well as the design at stake. While, as suggested above, the design teams within research centers dedicated to Values in Design typically include such scholars by default, realizing this goal in real-world design contexts remains a challenge. When approaching a technical design team outside the laboratory, it remains unclear who exactly is envisaged for assuming the role of who we call the "values advocate," and how one would formulate the criteria for this individual. Choosing either a social scientist or a philosopher to complement the design team, may lead to different outcomes; where social scientists have stronger training in empirical research and would therefore be better suited for keeping the empirical ball in play, philosophers and ethicists are generally better trained in analytical skills and may therefore be inclined to pay more intention to the conceptual ball, e.g. the clarification and normative determination of values.

The ideal scenario would be to have both types of advocates on the design team, but real-world practicalities often make this difficult. It would not be unusual for designers and engineers to be wary, if not openly hostile, to having someone whose purpose is to advocate for human values thrust onto their design teams. Attempting to add *two* such outsiders could likely spark more than double the resistance. Much, of course, depends on the particular design context the values advocate is trying to enter. With Web search engines, there is hope

that their origins in academia might make them more amenable to accepting multiple social scientists or ethicists among their ranks.[103]

A further consideration is the extent to which the values advocate must possess relevant engineering knowledge and expertise with regard to the aims and practical possibilities of the technology under consideration. The successful operationalization of values within a particular design context requires specific knowledge of a technology's design scope and limitations. Possessing the relevant technical expertise is also often necessary for acceptance as a member of a technical design team outside the walls of the academic laboratory. However, economic constraints of real-world design contexts might preclude the possibility of 'on-the-job' training for the values advocate, leaving a significant challenge for social scientists or philosophers in the role of the values advocate to obtain the requisite technical knowledge prior to engagement with the design community.[104] Gaining the necessary technical expertise in order to engage with Web search engine designers is particularly difficult given that little is known of exactly how search algorithms and related processes operate. While some of the foundational academic research is published, more recent glimpses into the inner workings of search engines are rarely provided,[105] limiting the ability for outside advocates to gain in-depth technical proficiency.

If these challenges can be overcome, we envision an array of roles that the values advocate can take within a particular design context. While rarely clear-cut, we can isolate three distinctive roles along a broad continuum of possible integration within design teams: the authoritarian role, the supporting role, and the collaborative role.[106] In the first role, the values advocate is regarded as authoritative in the ethical and value considerations at hand, implementing value-conscious decisions using a top-down strategy. Such a role is often confrontational, where pressure is exerted (both internally and externally) to

[103] Many of the first Web search engines, including Google and Yahoo!, originated as research projects at university computer or information science departments.

[104] A possible solution to both this challenge of gaining technical proficiency, and the previously-discussed concern of adding an outside social scientist or ethicist to the design team would be to foster and develop the ethical competencies of technical members already on design teams. Ensuring technically-trained engineers or computer scientists have the necessary background in ethics and values remains a challenge too complex to be directly addressed in this paper.

[105] A recent New York Times article was notable for its supposed "inside access" to Google's inner workings, but ultimately provided few details (see Hansell: 2007).

[106] These roles are by no means exhaustive or mutually exclusive, but they are three archetypes that are useful for conceptualizing the varied role of the values advocate.

ensure proper attention to values. In addition, in this scenario (attention to) values is warranted solely by the expertise of the acquired values advocate; this seems a heavy responsibility. A second role, the supporting role, is a less aggressive accompaniment of the design process, raising awareness at moments where value choices are being made and pointing to possible alternatives without advocating the one or the other. An advocate in this role is often merely advisory, and might have little more than token input on design decisions. Finally, the collaborative role enables the values advocate to fully participate as an equal on the design team, provide ongoing support to team members, but also assert necessary authority when it comes to making value choices. By providing insight to the complexity and delicacy of value choices basing on theoretical knowledge as well as acquired practical expertise, the values advocate in her role as collaborator is able to educate the other members of the design team (and possibly other stakeholders) and to strongly promote certain choices over others where necessary. We argue the collaborative role to be the preferred positioning of the values advocate within a design team, taking into account shared responsibility for value choices (whereas the first example of an authoritative role might be too heavy-handed) and a proactive stance of the values advocate (whereas the example of a supporting role might be too passive).

Recent efforts by advocacy groups concerned about Web search privacy help illustrate the strengths and weaknesses of these various roles. Privacy International, for example, took an authoritarian approach with its report 'A Race to the Bottom' (2007) strongly criticizing the privacy practices of major Internet companies, especially Google.[107] Alternatively, the Center for Democracy and Technology (2007) recently issued a report detailing how the largest Internet search companies have begun to aggressively compete with one another to offer stronger privacy protections to their customers. This study was written off, however, by those who criticized CDT for being financially supported by some of the very search engines in the report (see Singel: 2007). CDT's supporting role with the search engine companies essentially diluted the effectiveness of their advocacy. We argue that a middle-ground between these two approaches would place value advocates in a strong collaborative role to work with Web search engine providers. As a collaborator, the values advocate would be empowered, for example, to take the initiative to convene the leading search engine companies in a non-hostile environment to discuss, debate if necessary,

[107] Industry experts criticized the aggressive stance of the report as poorly researched, non-comprehensive, and biased (see Sullivan: 2007).

and decide how to move forward to design future Web search products in value-conscious ways.

In any of the three possible roles, it is the challenge for the values advocate to support all value choices with the necessary normative justification. The acceptance of a values advocate in the design team as holder of particular expertise is dependent on the perception that the value choices and positions held by the advocate are rationally and theoretically justified. We think any choice in the design process might potentially have moral import. It is the task of the values advocate to make these as explicit as possible. In some cases the values advocate may also need to justify his or her own presence in the design team, and why designers, or anyone for that matter, should be concerned with values of moral import in the first place.

6.4.4. Confronting competing values

Perhaps the most apparent challenge of applying Values In Design principles outside the academic laboratory is the inevitability of confronting competing values within varied design contexts. Rather than benefiting from working within the academic sphere committed to the primacy of designing for moral values, design contexts outside academia often include stakeholders whose goals might come into conflict with the protection of these values.

For example, Howe and Nissenbaum's TrackMeNot Web browser extension was developed using Mozilla's open-source browser application framework and posted to a website for download by the user and development community. As an academic project, little concern was necessary for production or distribution costs, advertising, profits, or other factors typically in play in commercial software development contexts. In contrast, the attempts at pragmatic engagement in design contexts outside academia by Zimmer and Manders-Huits exposed how the pressures of the marketplace – with its focus on instrumental values of efficiency, productivity and profitability – might result in hostility to the privileging of moral values in the design process. As Agre and Harbs (1994: 84) warn, standard-setting processes often "embed a wide variety of political agendas" and the process of developing those standards will be "contested along a variety of fronts by various parties." The same can be said for technical design overall; engaging in Values in Design within real-world design

contexts will require the ability to negotiate such instances of conflicting values.[108]

We claim that, for a fully engaged and pragmatic application of Values In Design to be successful, it must ensure that values are not only discovered and clarified, but also consciously and deliberately built into design, even if such embedding of values conflicts with other design objectives.[109] This questions the supremacy of some values over others (hierarchy of values) and the way value choices in design are dealt with in light of supporting value systems or political and ethical views. While some embrace a more cultural constructivist approach where values are taken as perceived by the majority of the people, we take a more normative position. Here, ethics provides grounds for identifying values and arguing for the enclosure or the supremacy of certain values over others, possibly against the (uncritical) majority opinion. We argue that, for morality to be designed into technical systems and institutions, value choices should be based on well-considered ethical judgments, coherent with our attitude towards how we think the world is best served and structured from a moral perspective. This involves a critical attitude towards certain aims of research and design that are often taken for granted, such as maximum efficiency, maximum profitability, and so on.

When dealing with a commercially operated Web search engine, for example, profitability and shareholder value will often come into conflict with attempts to design to support moral values. Specific to the collection of search query data, the search providers claim that they need such information to make their systems more efficient, and also to improve the usability and relevance of their results. The pursuit of these values come into direct conflict of ensuring user privacy, or a user's ability to search and use information free from surveillance or oversight. Applying ethical theory to overcome this value conflict provides

---

[108] To their credit, Flanagan and her colleagues make it clear in their discussion of the RAPUNSEL project that such projects, when pursued purely within academic contexts, fail to "address all factors that, for example, a commercial project would be obliged to address". Of course, Camp, Friedman, Flanagan and their collaborators likely acknowledge the privileged position of designing their products in the relatively non-hostile environment of the academic laboratory. Our intent is not to discredit their accomplishments, but merely to show that attention must be paid to the value conflicts that are inevitable to occur when engaged in Value-Conscious Design in traditional design contexts.

[109] This is why we prefer to speak of "Value-Conscious Design" as a normative interpretation of the VID framework (see also Manders-Huits and Zimmer: 2009, Manders-Huits: forthcoming).

different views. From a deontological perspective for instance, we can hold companies such as Google to their pledge "Don't be evil," and build the case that they have a moral obligation to protect user privacy to its fullest extent. However, there is also the option of a more consequentialist attempt to resolve this value conflict, where Google frequently argues that users will benefit from personalization of results and advertising, and any potential privacy threats are, on the whole, outweighed by these benefits.

This illustrates the complexity inherent in value conflicts, and the variety of arguments that can be crafted to try to resolve them to benefit one position over another. Our challenge, then, is to craft a strategy for pragmatic engagement to push their resolution in favor of value-conscious design of search engines to protect user privacy. We need to work towards confirming the first proposition (that search engines have a moral obligation to protect user privacy), while also confronting the second proposition (that users will not necessarily benefit from personalization, and any benefits do not automatically outweigh the potential harms).

6.5. Conclusion

We have identified three key challenges that must be addressed if scholars committed to Values In Design are to be successful in pragmatically engaging within technical design communities in support of the advancement of human and moral values. First, we confront the broad philosophical challenge of providing sufficient justification of the value and moral framework we embrace when engaging in VID. Second, the role of the values advocate must be both clearly defined and justified to fulfill its role as an essential component of technical design teams. Finally, we must find strategies to successfully resolve the inevitable value conflicts when engaging with design communities outside of academia. By viewing these challenges in the context of Web search query retention, we have brought to light some of the unique problems that emerge when engaging with real-world technical design communities.

Our goal with this chapter is not to discredit the accomplishments of the existing Values In Design efforts, but merely to show that pragmatic challenges remain. Moreover, like many academic probes, this chapter presents more questions than answers. By bringing focus to these challenges, however, we hope to spark new critical reflection within the Values In Design community (among which we include ourselves) of how to best engage with real-world

design communities to proactively influence the design of emerging technologies to take account of ethical and human values and help refine the Values In Design methodologies to increase the chances for success of future pragmatic engagements with design communities.

6.6. References

Ackerman, E. (2006, August 19) What do Google, Yahoo, AOL and Microsoft's MSN know about you? San Jose Mercury News.

Agre, P. and Harbs, C. (1994) 'Social Choice about Privacy: Intelligent Vehicle-highway Systems in the United States', in: *Information Technology and People* 7 (4): 63-90.

Associated Press. (2007) EU data privacy officers launch investigation into Google's Internet search engine. International Herald Tribune, Retrieved July 28, 2007, from http://www.iht.com/articles/ap/2007/05/25/business/EU-FIN-EU-Google-Privacy-Probe.php.

Barbaro, M. and Zeller Jr., T. (2006, August 9) A Face Is Exposed for AOL Searcher No. 4417749, The New York Times, A1.

Battelle, J. (2006a) More On What Google (and Probably A Lot of Others) Know, Searchblog, Retrieved May 16, 2006, from http://battellemedia.com/archives/002283.php.

----- (2006b) What Info Does Google Keep?, Searchblog, Retrieved May 16, 2006, from http://battellemedia.com/archives/002272.php.

Berg, M. (1998) 'The Politics of Technology: On Bringing Social Theory into Technological Design', in: *Science, Technology, and Human Values* 23 (4): 456-490.

Boneh, D., Boyen, X., and Shacham, H. (2004) 'Short group signatures', in: *Proceedings of Crypto* 4: 41-55.

Brin, S., and Page, L. (1998) 'The Anatomy of a Large-Scale Hypertextual Web Search Engine', in: *WWW7 / Computer Networks* 30 (1-7): 107-117.

Camp, L. J. (2006) Reliable, Usable Signaling to Defeat Masquerade Attacks, Retrieved September 23, 2006, from http://www.ljean.com/files/NetTrustEcon.pdf.

----- (n.d.) Design for Values, Design for Trust. Retrieved September 20, 2006, from http://www.ljean.com/design.html.

Camp, L. J., Friedman, A., and Genkina, A. (n.d.) Embedding Trust via Social Context in Virtual Spaces, Retrieved September 23, 2006, from http://www.ljean.com/files/NetTrust.pdf.

Center for Democracy and Technology. (2007) Search Privacy Practices: A Work In Progress, Retrieved August 8, 2007, from http://www.cdt.org/privacy/20070808searchprivacy.pdf.

Electronic Frontier Foundation. (2007) Privacy and Search Engines. Retrieved July 28, 2007, from http://www.eff.org/Privacy/search/.

Flanagan, M., Howe, D. and Nissenbaum, H. (2008) 'Embodying Values in Technology: Theory and Practice', in: van den Hoven, J. and Weckert, J. (eds.), *Information Technology and Moral Philosophy*, Cambridge University Press: 322-353.

Friedman, B. (1997) *Human Values and the Design of Computer Technology*, Cambridge University Press, New York.

----- (1999) Value-Sensitive Design: A Research Agenda for Information Technology, National Science Foundation, Contract No: SBR-9729633.

Friedman, B., Howe, D., and Felten, E. (2002) 'Informed consent in the Mozilla browser: Implementing value-sensitive design', in: *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*.

Friedman, B., Kahn, P., and Borning, A. (2002) Value Sensitive Design: Theory and Methods, (Technical Report 02-12-01).

Glasner, J. (2005) What Search Sites Know About You. Wired News, Retrieved August 2, 2006, from http://www.wired.com/news/privacy/0,1848,67062,00.html.

Google. (2005) Google Privacy Policy, Retrieved May 3, 2006, from http://www.google.com/privacypolicy.html.

Hafner, K., and Richtel, M. (2006, January 20) Google resists U.S. subpoena of search data, The New York Times A1, C4.

Hansell, S. (2005, September 26) Microsoft plans to sell search ads of its own, The New York Times, C1, C8.

----- (2006, August 8) AOL Removes Search Data On Vast Group Of Web Users, The New York Times, C4.

----- (2007) Google Keeps Tweaking its Search Engine, New York Times, Retrieved July 28, 2007, from http://www.nytimes.com/2007/06/03/business/yourmoney/03google.html.

Howe, D., and Nissenbaum, H. (2006) TrackMeNot, Retrieved August 27, 2006, from http://mrl.nyu.edu/~dhowe/TrackMeNot.

IAC Search & Media (2005) Privacy Policy for Ask.com, Retrieved January 6, 2007, from http://sp.ask.com/en/docs/about/privacy.shtml.

Johnson, D., and Nissenbaum, H. (eds.) (1995) *Computers, Ethics, and Social Values,* Englewood Cliffs, N.J, Prentice Hall.

Latour, B. (1992) 'Where are the missing masses? The sociology of a few mundane artifacts', in: Bijker, W. and Law, J. (eds.), *Shaping technology/building society*: S*tudies in sociotechnical change*, MIT Press, Cambridge MA: 225-258.

Lohr, S. (2007) Google Deal Said to Bring U.S. Scrutiny, The New York, Times Retrieved July 27, 2007, from http://www.nytimes.com/2007/05/29/technology/29antitrust.html.

MacKenzie, D. A., and Wajcman, J. (eds.) (1985) *The Social Shaping of Technology: How the Refrigerator got its Hum*, Open University Press, Philadelphia.

Manders-Huits, N. (forthcoming) 'What Values in Design? The Challenge of Incorporating Moral Values into Design', in: *Science and Engineering Ethics*. Online first http://www.springerlink.com/content/d28305n23v9715t3/.

Manders-Huits, N. and Zimmer, M. (2009) 'Values and Pragmatic Action: The Challenges of Introducing Ethical Intelligence in Technical Design Communities', in: *International Review of Information Ethics* 10: 37-44.

McCullagh, D. (2006) AOL's disturbing glimpse into users' lives, CNET News.com, Retrieved December 3, 2006, from http://news.com.com/AOLs+disturbing+glimpse+into+users+lives/2100-1030_3-6103098.html?tag=st.num.

Mintz, H. (2006) Feds after Google data: Records sought in U.S. quest to revive porn law, San Jose Mercury News, Retrieved January 19, 2006, from http://www.siliconvalley.com/mld/siliconvalley/13657386.htm.

Mitcham, C. (ed.) (2005) *Encyclopedia of Science, Technology, and Ethics*, Macmillan Reference, Detroit.

Moor, J. (1985) 'What is computer ethics?', in: *Metaphilosophy* 16: 266-275.

Mumford, L. (1964) 'Authoritarian and Democratic Technics', in: *Technology and Culture*, 5 (1): 1-8.

Nielsen, J. (1993) *Usability Engineering*, Academic Press, Boston.

Nissenbaum, H. (2001) 'How computer systems embody values', in: *IEEE Computer* 34 (3): 118-120.

Nissenbaum, H., Flanagan, M., and Howe, D. (2005) 'Values at Play: Design Tradeoffs in Socially-Oriented Game Design', in: Proceedings of CHI 2005: 751-760.

Norman, D. A. (1990) *The Design of Everyday Things*, Doubleday, New York.

Page, L., Brin, S., Motwani, R., and Winograd, T. (1998) The pagerank citation ranking: Bringing order to the web, Retrieved January 12, 2007, from http://dbpubs.stanford.edu/pub/1999-66.

Pitkow, J., Schütze, H., Cass, T., Turnbull, D., Edmonds, A., and Adar, E. (2002) 'Personalized search', in: *Communications of the ACM* 45 (9): 50-55.

Postman, N. (1992) *Technopoly: The Surrender of Culture to Technology*, Vintage Books, New York.

Privacy International. (2007) A Race to the Bottom: Privacy Ranking of Internet Service Companies, Retrieved July 10, 2007, from http://www.privacyinternational.org/article.shtml?cmd%5B347%5D=x-347-553961.

Raskin, J. (2000) *The Humane Interface: New Directions for Designing Interactive Systems*, Addison-Wesley, Reading.

Reiman, J. (1995) 'Driving to the panopticon: A philosophical exploration of the risks to privacy posed by the highway technology of the future', in: *Santa Clara Computer and High Technology Law Journal* 11 (1): 27-44.

Sclove, R. (1995) *Democracy and Technology*, Guilford, New York.

Shrader-Frechette, K., and Westra, L. (eds.) (1997) *Technology and Values*, Rowman & Littlefield, Lanham MD.

Singel, R. (2007) Industry-funded Group and a Compliant THREAT LEVEL Working To Derail Google-Doubleclick Probe, Privacy Group Charges, Threat Level, Retrieved August 10, 2007, from http://blog.wired.com/27bstroke6/2007/08/industry-funded.html.

Sullivan, D. (2007) Google Bad on Privacy? Maybe It's Privacy International's Report That Sucks, Search Engine Land, Retrieved July 29, 2007, from http://searchengineland.com/070610-100246.php.

Tavani, H. T. (2004) *Ethics And Technology: Ethical Issues In An Age Of Information And Communication Technology*, Wiley, Hoboken NJ.

Teevan, J., Dumais, S. T., and Horvitz, E. (2005) 'Personalizing search via automated analysis of interests and activities', in: *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*: 449-456.

Van Den Hoven, M. J. (2005) 'Design For Values And Values For Design', in: *Journal of the Australian Computer Society* 7 (2): 4-7.

Winner, L. (1980) 'Do Artifacts have Politics?', in: *Daedalus* 109 (1): 121-136.

Yahoo! (2006) Yahoo! Privacy Policy, Retrieved January 6, 2007, from http://info.yahoo.com/privacy/us/yahoo/details.html.

Epilogue

The inquiry of "Designing for Moral Identity" belongs to the research field of 'ethics of technology.' Scholarly attention to ethics of technology, especially to 'values in design' is of a recent date. Prior to this development, technology was usually portrayed as 'context-free' and 'value-neutral.' Despite of the fact that questions about usability or functionality of technology in practice often lead to discussions of ethical values, e.g. equality, autonomy, or transparency, the focus of solving problems concerning technology has mainly been on developing the technology instead of investigating its relation to ethical (value) frameworks. Recently, the successful application of technology is increasingly seen as *dependent* on its capacity to accommodate human values and moral considerations. Simultaneously, ethics has also evolved from a principally theoretical enterprise to offering applied analyses regarding technology, economy, institutional and legal frameworks and incentive structures that constitute the moral situation. Moreover, ethicists realize that to bring about real and desirable moral changes in the world, we need to design systems, institutions, infrastructures in accordance with moral values. Hence the new challenge of explicitly addressing ethical values in technology and proactively giving direction to the interaction between ethical values and technology in design. This is what I refer to as Value-Conscious Design.

In this thesis I encourage the taking of a moral attitude towards designing identity related information technologies. The moral attitude I argue comprises epistemic modesty, diligence, and restraint. This results in basic principles for design such as data minimization and purpose specification, i.e. defining the objectives for each particular technology and its use, redefining these objectives if necessary and keeping to them. On a more practical level, there are numerous ways of practically translating and implementing these directives into requirements and specifications for design. In turn these can be satisfied by developing value-conscious technical tools and applications.

In this thesis I put forward the ideal of accommodating moral values in design; why we should be interested in this is self-evident. Therefore the way forward is the further integration of ethics and technology (design). This implies refining the methodology for implementing values in design and measuring the success of applying such methodology. Especially with regard to identity related information technologies, the main object of study in this research, ethics

deserves attention in order to connect with the way users experience technology as regards identity from a moral perspective. In future research, designers should start from the moral attitude towards design as I put forward in this thesis and incorporate the associated moral considerations concerning identity in their designs. Above all this topic deserves attention in relation to the new generation of identity- and profile-based information technology applications.

I believe we should not get carried away by the innumerable opportunities information technology provides us with now and in the future, but we should try to design identity related information technology while paying attention to the relevant values concerning identity as a morally significant category for users. Value-Conscious Design provides us with the opportunity of making the world a better place for people who have ideas about who they want to be and about how they want to be known.

Summary

This thesis sets out to identify moral considerations for informing the design of Identity Management (IDM) and profiling technologies. I refer to these technologies as "identity related information technologies." This thesis is a collection of separately published papers and should not to be regarded as a set of chronologically ordered or logically consecutive contributions. The papers are all related to the overall research aim of this PhD-project, namely to contribute to the value-conscious design of identity related information technologies.

A relevant question concerns the impact of the ethical considerations that I describe in this thesis. The obvious harmful consequences associated with identity related technologies involve identity theft, privacy infringement, unlawful use of identity related information. I have chosen not to focus primarily on these issues, since they are already well-known and (also) part of the legal and policy domain. What I am interested in is a less straightforward type of harm. It involves the origin, justification and understanding of moral feelings of uneasiness with and intuitive resistance against the use of a person's identities and profiles. I argue that in particular circumstances and under certain conditions these negative moral feelings about the utilization of identity related information are warranted. As I describe in chapter 3, this type of harm, i.e. harm to moral identity resulting from information technology design (and use), is for the most part hardly noticeable nor tangible, at the moments it takes place. Its harmful effects are best understood in the accumulative sense, like some forms of pollution of the environment or exhaustion of common pool resources. For example, the coupling of multiple databases containing information regarding the same person may bring about (new) identity related information that is sensitive in this particular context. Moreover, it can then lead to undesired consequences that could not have been foreseen prior to the accumulation. Notwithstanding the intangible nature and elusiveness of these concerns regarding our identity related data, they do in fact influence our behavior. Once we are aware of the fact that information about us may turn up in future circumstances, and there and then may start to affect our lives, we may become far more reserved in acting and sharing information (especially online).

This research explores the relation between ethics and technology, more specifically, the relation between identity in a moral sense and IDM and profiling technologies. The first part (chapters 2-4) provides a conceptual

analysis of relevant philosophical concepts and tools, whereas the latter two chapters (5-6) focus on the challenges of integrating moral theories, values and concepts in technology design. In chapter 2, I introduce the notions of "moral" and "practical" identity. I argue in this chapter that information technologies impose a practical, i.e. an administrative, forensic, conception of identity and identification that is distinctly different from the sense of identity that a moral conception of self-presentation and identity management requires. Looking at identity through the looking glass of IDM and profiling technologies unduly limits persons with respect to their self-presentation and management of their moral identity. In chapter 3 I describe the epistemic gains of IDM, as well as the type of harm that is central to this thesis. The 'informational' structuring of public administration and monitoring of citizens implies a type of harm which seems nearly invisible and negligible in each particular case, but becomes apparent in a collective, accumulated fashion. This, so I argue, applies to the collecting and processing of personal - or identity related - data. It is what I call "accumulative informational harm." In chapter 4 I develop one of the main moral requirements for 'Designing for Moral Identity,' i.e. the requirement of "moral identification." I explore the obligations we have to persons when we represent them in databases, when we record and use their identities, and when we interact with them on the basis of these representations. The chapter articulates the main moral reasons for imposing constraints upon the development and use of IDM technologies by persons other than those whose identities are at stake. In chapter 5 I explore if and how moral value requirements, such as the requirement of moral identification can be designed into information technology. The focus here is on the criteria of adequacy for an approach or methodology for integrating moral values into the design of technologies in a way that joins in with an analytical perspective on ethics of technology. The suitability of VSD as the prime candidate for implementing moral values in design is considered from a conceptual, analytical, normative perspective. In chapter 6 I focus on the pragmatic challenges of engaging in the implementation of moral values in design, especially in real-world design settings. In that chapter I describe two attempts to practically engage with (technical) design communities in order to influence the design of these technologies in value-conscious ways, revealing discouraging results. Learning from these failed attempts, this chapter identifies several key challenges of pragmatic engagement with technical design communities.

This research is a contribution to the overall aim of designing our identity related information technologies in a value-conscious way, i.e. a way which accommodates the values concerning our identity as a morally significant phenomenon. As I explain in chapters 5 and 6, the basic thought of this approach, and similar approaches, is to tease out norms and values associated with and informing or driving the application and implementation of the technology in question. Uncovering relevant norms and values, or the ethical framework, is not a fixed part of the Value-Conscious Design process, but needs to be consciously and deliberately given substance in each case. This includes any presuppositions and assumptions held by designers and stakeholders. Explicating and investigating these norms and values - in the case of this research concerning (moral) identity perception and identification - is necessary for realizing a value-conscious design.

Value-consciously designing technology thus entails:

1. An open **attitude** of designers towards ethical values and the desire to design the technology conscious in light of (shared deliberated) ethical values;
2. Favorable **conditions** regarding the flexibility both of the design team and the technology in question, e.g. the possibility for including a values advocate in the design team and the possibility for changing the design in accordance with deliberated values;
3. The explication and deliberate selection of an ethical framework, i.e. the realization of situation-specific **norms and values**;
4. The translation of this framework into **technical requirements** and tools;
5. The **monitoring and assessment** of values associated with the technology under construction in relation to the selected framework.

In addition to the exploration of possible means to integrate philosophical conceptual work into design (chapters 5-6), this thesis also provides a conceptual analysis of moral identity perception in relation to identity management and profiling technologies (chapters 2-4). This results in the following ethical guidelines for designing identity related information technologies:

a. An attitude of modesty in epistemic claims regarding persons;
b. Enabling data subjects to endorse or identify with their acquired or imposed identities;
c. A restraint in generating, collecting and storing data without clear objectives;

d. The possibility of perusal of and participation in the construction and adjustment of available identities and profiles concerning oneself;
e. The active prevention of stereotyping and stigmatization by means of continuous assessment of available identities and profiles.

This thesis offers the necessary preliminary conceptual work in order to substantiate moral claims and considerations regarding identity in relation to IDM and Profiling technologies. These considerations can in turn be used for developing the actual technical requirements for a "Value-Sensitive," or "Value-Conscious," (cf. Ch 5) Design. There are numerous ways of practically translating and implementing these directives. To name a few, in chapter 3 I mention *identicons* as tools for indicating the data aggregation threshold concerning a user. Another possible practical implementation of the guidelines mentioned above are *notification* messages whenever identity related data concerning a user is demanded by a party, or whenever identity related data is exchanged between different parties. This way the user is actively engaged in the aggregation and exchange of his or her identity related data. Moreover, the user can decide whether he or she is satisfied with the amount of data collected and produced, whether he or she thinks certain information should be exchanged with a certain party, and whether he or she thinks certain identity related data is not (yet) sufficient and should be complemented. A third option for engaging the person whose identity is at stake is to create a tool for *endorsing* constructed identities and profiles.

In summary, this thesis hopefully contributes to the further integration of ethics and technology, with a particular focus on identity related information technologies such as identity management and profiling technologies. For one thing it clarifies the moral concept of identity and other relevant concepts such as identity related data in relation to identity related information technologies. It also presents a relevant conception of moral harm in addition to the more straightforward, tangible harms concerning identities and their data. For another it discusses the challenges of implementing ethical considerations into the design of (any) technology, both from a philosophical, methodological perspective, as from a practical perspective. Finally, this thesis puts forward a desirable moral attitude towards the design of identity related information technologies (e.g. identity management and profiling technologies). The attitude implies respect for persons with regard to epistemic claims concerning their identities, diligence in data exchange and mining, and restraint in data aggregation.

Samenvatting

Het doel van deze dissertatie is de morele overwegingen te identificeren die van belang zijn bij het ontwerp van technologieën voor identiteitsmanagement (IDM) en profieltechnologieën. Ik noem deze technologieën "identiteits-gerelateerde informatietechnologieën." Deze dissertatie is een verzameling van afzonderlijk gepubliceerde artikelen die niet als een reeks chronologisch geordende of logisch opeenvolgende bijdragen moeten worden gezien. Alle artikelen zijn gerelateerd aan de algemene doelstelling van dit promotie-onderzoek, namelijk een bijdrage te leveren aan het waardebewust ontwerpen van identiteitsgerelateerde informatietechnologieën.

Een relevante vraag betreft de mogelijke impact van de ethische overwegingen die ik in deze dissertatie beschrijf. De duidelijk schadelijke gevolgen die met identiteitsgerelateerde informatietechnologieën worden geassocieerd zijn identiteitsdiefstal, inbreuk op privacy en onrechtmatig gebruik van identiteitsgegevens. Ik heb ervoor gekozen om me niet hoofdzakelijk op deze gevolgen te richten, omdat deze al bekend zijn en (ook) deel uit maken van het juridische en politieke domein. Ik ben juist geïnteresseerd in minder voor de hand liggende soorten schade, zoals de herkomst, rechtvaardiging en het begrijpen van morele gevoelens van onbehagen over en intuïtief verzet tegen het gebruik van iemands identiteitsgegevens en profielen. Ik beweer dat onder bepaalde omstandigheden en onder bepaalde voorwaarden deze negatieve morele gevoelens over het gebruik van identiteitsgerelateerde informatie gegrond zijn. Zoals ik in hoofdstuk 3 beschrijf is dit type schade, d.w.z. schade aan de morele identiteit door het ontwerp (en gebruik) van informatie-technologie, over het algemeen nauwelijks waarneembaar of tastbaar op het moment dat deze plaatsvindt. De schadelijke effecten zijn pas duidelijk bij accumulatie, net als bij sommige vormen van milieuvervuiling of uitputting van gemeenschappelijke bronnen. Zo kan het koppelen van meerdere bestanden met informatie over dezelfde persoon, leiden tot (nieuwe) identiteitsgerelateerde informatie die in een bepaalde context gevoelig kan zijn. Dit kan ongewenste gevolgen hebben die voorafgaand aan de accumulatie van informatie niet waren voorzien. Ook al zijn deze zorgen over onze identiteitsgerelateerde informatie niet tastbaar en ongrijpbaar, ze beïnvloeden ons gedrag wel degelijk. Als we ons er eenmaal van bewust zijn dat informatie over onszelf in de toekomst onder andere omstandigheden gebruikt kan worden en van invloed kan zijn op ons

leven, zullen we voorzichtiger zijn in ons handelen en in het delen van informatie (met name online).

Dit onderzoek gaat dieper in op de relatie tussen ethiek en technologie, in dit geval de relatie tussen identiteit in moreel opzicht en identiteitsgerelateerde informatietechnologieën. In het eerste deel (hoofdstukken 2-4) wordt een conceptuele analyse van relevante filosofische concepten en instrumenten beschreven. De laatste twee hoofdstukken (5-6) zijn gewijd aan de ingewikkelde opgave om morele theorieën, waarden en concepten te integreren in het technologisch ontwerp. In hoofdstuk 2 introduceer ik de begrippen "morele" en "praktische" identiteit. In dit hoofdstuk stel ik dat informatietechnologieën een praktisch, d.w.z. een administratief en forensisch begrip van identiteit en identificatie opleggen, dat wezenlijk verschilt van het besef van identiteit dat een moreel begrip van zelfpresentatie en identiteitsmanagement vereist. Identiteit bezien door de bril van IDM en profieltechnologieën beperkt mensen in hoge mate in hun zelfpresentatie en beheer van hun morele identiteit. In hoofdstuk 3 beschrijf ik de epistemologische voordelen van IDM, alsook het soort schade dat centraal staat in deze dissertatie. De 'informationele' structurering van het openbaar bestuur en controle van burgers brengt een type schade met zich mee dat in afzonderlijke gevallen nagenoeg onzichtbaar en te verwaarlozen lijkt, maar bij accumulatie duidelijk wordt. Dit, zo betoog ik, geldt bij het verzamelen en verwerken van persoonlijke of identiteitsgerelateerde gegevens. Dit noem ik "accumulatieve informationele schade." In hoofdstuk 4 beschrijf ik een van de belangrijkste morele vereisten voor het 'ontwerpen voor morele identiteit,' die van "morele identificatie." Ik ga dieper in op de verplichtingen die we hebben ten aanzien van personen wanneer we hun gegevens opslaan, hun identiteit vastleggen en gebruiken, en hen aanspreken op basis van deze representaties. In dit hoofdstuk beschrijf ik de belangrijkste morele redenen voor het opleggen van beperkingen bij het ontwikkelen en toepassen van IDM-technologieën door anderen dan diegenen op wie deze identiteiten betrekking hebben. In hoofdstuk 5 onderzoek ik of en hoe vereisten voor morele waarden, zoals de vereiste van morele identificatie, geïntegreerd kunnen worden in het ontwerp van de informatietechnologie. De nadruk ligt hierbij op de adequaatheidscriteria voor een aanpak of methodologie voor het integreren van morele waarden in het ontwerp van technologieën op een manier die aansluit bij een analytisch perspectief op de ethiek van technologie. De geschiktheid van "Value-Sensitive Design" (VSD) als de beste manier voor het implementeren van morele waarden in het ontwerp van technologieën wordt beschouwd vanuit een conceptueel,

analytisch, normatief perspectief. In hoofdstuk 6 ligt de nadruk op de pragmatische uitdagingen bij de implementatie van morele waarden in het ontwerp, vooral in realistische ontwerpscenario's. In dit hoofdstuk beschrijf ik twee pogingen om ontwerpers te betrekken bij het implementeren van morele waarden in het ontwerp van technologieën, met teleurstellende resultaten. Uit deze mislukte pogingen komen enkele belangrijke hindernissen naar voren bij de praktische samenwerking met de technische gemeenschap.

Dit onderzoek is een bijdrage aan de algemene doelstelling om identiteitsgerelateerde informatietechnologieën te ontwerpen op een waardebewuste manier, d.w.z. een manier die recht doet aan de waarden met betrekking tot onze identiteit als een moreel significant fenomeen. Zoals ik uitleg in hoofdstukken 5 en 6, vormt het distilleren van de normen en waarden die verband houden met en ten grondslag liggen aan de toepassing en implementatie van de desbetreffende technologie de basis van deze benadering en soortgelijke benaderingen. Het achterhalen van de specifieke relevante normen en waarden, ofwel het ethische raamwerk, moet per geval bewust en weloverwogen worden vormgegeven. Dit geldt tevens voor de vooronderstellingen en aannames van ontwerpers en stakeholders. Het formuleren en achterhalen van deze normen en waarden – in het geval van dit onderzoek met betrekking tot (morele) identiteitsperceptie en -identificatie – is noodzakelijk voor een waardebewust ontwerp.

Waardebewust ontwerpen vraagt derhalve:

1. Een open **attitude** van ontwerpers ten opzichte van ethische waarden en de wens de technologie bewust te ontwerpen met het oog op (gezamenlijk overeengekomen) ethische waarden;
2. **Gunstige omstandigheden** met betrekking tot de flexibiliteit van zowel het ontwerpteam als de technologie in kwestie, bijvoorbeeld de mogelijkheid een teamlid te belasten met het bewaken van de afspraken en de mogelijkheid het ontwerp aan te passen aan de overeengekomen waarden;
3. Het formuleren en opstellen van een ethisch raamwerk, d.w.z. het realiseren van situatiespecifieke **normen en waarden**;
4. De vertaling van dit raamwerk in **technische vereisten** en instrumenten;
5. Het **bewaken en beoordelen** van waarden met betrekking tot de technologie in kwestie in relatie tot het opgestelde raamwerk.

Naast het onderzoeken van de mogelijkheden om filosofisch conceptueel werk te integreren in het ontwerp (hoofdstukken 5-6), biedt deze dissertatie ook een conceptuele analyse van de perceptie van morele identiteit in relatie tot identiteitsmanagement- en profieltechnologieën (hoofdstukken 2-4). Dit resulteert in de volgende ethische richtlijnen voor het ontwerpen van identiteitsgerelateerde informatietechnologieën:

a. Een bescheiden opstelling bij kennisclaims betreffende personen;
b. De mogelijkheid voor degene van wie gegevens worden vastgelegd om zijn of haar verkregen of opgelegde identiteiten goed te keuren of zich ermee te identificeren;
c. Terughoudendheid bij het genereren, verzamelen en opslaan van gegevens zonder duidelijk oogmerk;
d. De mogelijkheid voor een persoon om de beschikbare identiteiten en profielen van zichzelf in te zien of mee te werken aan het opstellen en aanpassen ervan;
e. Actieve preventie van stereotypering en stigmatisering door voortdurende beoordeling van de beschikbare identiteiten en profielen.

Deze dissertatie biedt het noodzakelijke voorbereidende conceptuele werk om morele beweringen en overwegingen omtrent identiteit met betrekking tot IDM en profieltechnologieën te kunnen staven. De overwegingen kunnen vervolgens worden gebruikt bij de ontwikkeling van de uiteindelijke technische vereisten voor een 'waardegevoelig' of 'waardebewust' (cf. H5) ontwerp. Er zijn talloze manieren voor de praktische vertaling en toepassing van deze richtlijnen. Zo noem ik in hoofdstuk 3 *identicons* als middel om aggregatie van gegevens over een gebruiker te beperken. Een andere mogelijke praktische toepassing van de hierboven genoemde richtlijnen zijn *meldingen* die worden verstuurd telkens wanneer identiteitsgerelateerde gegevens betreffende een gebruiker door een partij worden opgevraagd of tussen verschillende partijen worden uitgewisseld. Op deze manier is de gebruiker actief betrokken bij het verzamelen of uitwisselen van zijn of haar identiteitsgerelateerde gegevens. Bovendien kan de gebruiker aangeven of hij of zij tevreden is met de hoeveelheid gegevens die wordt verzameld en geproduceerd, en of bepaalde informatie aan een bepaalde partij moet worden doorgegeven, en of bepaalde identiteitsgerelateerde gegevens (nog) niet compleet zijn en moeten worden aangevuld. Een derde mogelijkheid

om de persoon in kwestie erbij te betrekken is de ontwikkeling van een tool voor het *goedkeuren* van aangemaakte identiteiten en profielen.

Samengevat hoop ik dat deze dissertatie zal bijdragen aan een verdere integratie van ethiek en technologie, in het bijzonder identiteitsgerelateerde informatietechnologieën, zoals identiteitsmanagement- en profiel-technologieën. In de eerste plaats wordt een verklaring gegeven van het morele concept van identiteit en andere relevante concepten zoals identiteitsgerelateerde gegevens in relatie tot identiteitsgerelateerde informatietechnologieën. Tevens wordt een relevante omschrijving van morele schade gegeven in aanvulling op de meer bekende, tastbare schade met betrekking tot identiteiten en hun gegevens. Daarnaast worden de uitdagingen besproken van het implementeren van ethische overwegingen in het ontwerp van technologie, zowel uit een filosofisch, methodologisch, als een praktisch perspectief. Ten slotte oppert deze dissertatie een gewenste morele attitude ten aanzien van het ontwerp van identiteitsgerelateerde informatietechnologieën (bijv. identiteitsmanagement- en profiel-technologieën). Deze attitude houdt respect in voor personen met betrekking tot epistemische beweringen omtrent hun identiteiten, zorgvuldigheid bij het uitwisselen en verzamelen van gegevens en terughoudendheid bij het samenvoegen van gegevens.

About the Author

Noëmi Manders-Huits completed her PhD research on *Designing for Moral Identity in Information Technology* at the philosophy section of the department of Technology, Policy, and Management at Delft University of Technology, The Netherlands. Before this she studied philosophy and business administration at the Erasmus University in Rotterdam. She received her M.A. in (general) Philosophy in 2002 and graduated in 2004 as M.A. in Philosophy of Information Technology on a Master's thesis on *The Philosophical Semantics of Personal Data*. Her main research interest is in values in design and moral identity. Noëmi organized the three-day international conference on Ethics, Technology, and Identity in Delft/The Hague 2008. Since 2007 she is managing editor of the Springer-journal *Ethics and Information Technology*.

Simon Stevin (1548-1620)

'Wonder en is gheen Wonder'

This series in the philosophy of technology is named after the Dutch / Flemish natural philosopher, scientist and engineer Simon Stevin. He was an extraordinary versatile person. He published, among other things, on arithmetic, accounting, geometry, mechanics, hydrostatics, astronomy, theory of measurement, civil engineering, the theory of music, and civil citizenship. He wrote the very first treatise on logic in Dutch, which he considered to be a superior language for scientific purposes. The relation between theory and practice is a main topic in his work. In addition to his theoretical publications, he held a large number of patents, and was actively involved as an engineer in the building of windmills, harbors, and fortifications for the Dutch prince Maurits. He is famous for having constructed large sailing carriages.

Little is known about his personal life. He was probably born in 1548 in Bruges (Flanders) and went to Leiden in 1581, where he took up his studies at the university two years later. His work was published between 1581 and 1617. He was an early defender of the Copernican worldview, which did not make him popular in religious circles. He died in 1620, but the exact date and the place of his burial are unknown. Philosophically he was a pragmatic rationalist for whom every phenomenon, how-ever mysterious, ultimately had a scientific explanation. Hence his dictum 'Wonder is no Wonder,' which he used on the cover of several of his own books.

Simon Steven Series in the Ethics of Technology

Editors: Philip Brey, Peter Kroes and Anthonie Meijers

3TU. Centre for Ethics and Technology Publications

Dissertations:

Volume 1: Lotte Asveld, *Respect for Autonomy and Technologal Risks*, 2008.

Volume 2: Mechteld-Hanna Derksen, *Engineering Flesh. Towards Professional Responsibility for "Lived Bodies" in Tissue Engineering*, 2008.

Volume 3: Govert Valkenburg, *Politics by all means. An Enquiry into Technological Liberalism*, 2009.

Volume 4: Noëmi Manders-Huits, *Designing for Moral Identity in Information Technology*, 2010.