Broadening the Privacy Concept in the Digital Age: Adjusting rights?

Martijn van Otterlo, PhD
http://martijnvanotterlo.nl
mail@martijnvanotterlo.nl

Position paper prepared for the
Amnesty Strategic Investigations Symposium on Surveillance and Human Rights
(26th November 2014, Amsterdam)

I – Introduction

Privacy has gained a lot of attention recently. However, much is directed at the limited concept of private data and access to that data. Here I argue for a **much broader concept of privacy** and I sketch how new privacy threats are fundamental threats to how our society functions. As a start of a solution, I propose to explore ways to ensure our basic human rights in digital worlds. Our digital society has allowed powerful entities to grow unboundedly, and it is time to start thinking about how to restore the **power imbalance** that is now arising. I will first outline a broader view on privacy and surveillance I have developed in recent papers. The core consists of **smart algorithms** and the science of human behavior. Afterwards I will argue for a **universal declaration of human rights for digital worlds.**

II – An Algorithmic View on the Privacy Landscape

Privacy is typically equated with aspects of life that someone does not want others to see. In our digital world, privacy issues are much more complex and diverse; I like to characterize privacy in terms of *control-over-information*. In the following I describe three views on privacy, based on four fundamental notions: **data** and **measurement**, **generalization** and **prediction**, **action** and **manipulation**, and **feedback**.

A – Privacy: Who-can-see-what-information

The first type of control centers on ***access***. Much of the contemporary privacy debate is about what companies and governments know about us; what **data** they have. Framed as **big data** information is gathered everywhere, through various **measurements** coming from mobile phones, internet services, social media and public transport cards, to name just a few. Long before the infamous *"don't be evil"*-slogan by Google, Warner and Stone (1970, p146) warned us to "not be naive about it":

> *"Anyone who has entered into a hire-purchase transaction … should nowadays expect both the personal data he supplied in his application, and the information about his reliability in making the repayments, to be widely available."*

Information is available in digital form and privacy violations and abuse of data are common. Data is a commodity and is traded on a large scale. Control over information here is essentially about ***control of access,*** only guaranteed if only those (legal) persons who are "allowed" to see the data, can see it. On Facebook we can see a strange paradox: while getting fine-grained control over who can see which parts of a profile, we give Facebook itself total access to *all* our information (see also Otterlo 2013a).

B – Privacy: Profiles and Prediction

A second type of control arises when *artificially intelligent algorithms* (Nilsson, 2010) are employed, such as *machine learning and datamining*. These combine modern *statistical methods* with powerful *knowledge representation* languages to generate rich ***prediction models*** from data, to **generalize** and to generate *new* or *inferred* knowledge. Such models are based on information about many individuals, and they can be used to predict traits for individuals; see van Otterlo (2013) for an introduction. For example, models (probabilistically) predict whether I would buy a particular book, based on previous purchases and "similar" books. Models may contain a *generalized rule* "if a person is tall, it is more likely that it is a male", representing a typical *pattern* in the data. The rule may not predict well for *every* individual, but it surely predicts well *on average*. In addition, the rule may be used to predict (a possibly *unobserved* feature) "male" from (an *observed* feature) "tall".

Generating prediction models from data is essentially a form of complex **statistics**, which means that many *concerns* apply about conclusions drawn from data and about the amount of **bias** that went into the models. For individuals this is important –- what is being predicted, how are the models generated, and how accurate are they? – especially if your mortgage or service depends on it (cf. van Otterlo, 2013a). Schwartz et al. (2013) learn models based on *language use*: *"Our technique leverages what people say in social media to find distinctive words, phrases, and topics as functions of known attributes of people such as gender, age, location, or psychological characteristics."* The selection of which words and phrases to include is *data-driven* and comes from social media messages (e.g. Facebook status updates). Interestingly, many personality traits are statistically predictable from the use of particular words. Kosinski et al. (2013) use "Facebook like" behavior to similar models that predict similar features such as *religion*, *gender* and *sexual orientation*, with high accuracy.

C – Privacy: Behavioral Engineering and Experimentation

The third type of control over information deals with the **use** of data and prediction models for a purpose, such as surveillance or commercial profit. Warner and Stone (1970, p124) wrote:

> *"Give the administrator in government or business the use of an integrated national population file ... and you provide him with a powerful tool for interference in private lives, to manipulate, to sell more, to condition, to coerce."*

I have outlined (van Otterlo, 2014a, 2014b) how modern algorithms give rise to contexts governed by the principles of **behavioral engineering** as described in Walden Two by the psychologist B.F. Skinner. The idea is simple, yet very powerful: given a prediction model of individual behavior a company or government can exploit that model to **manipulate** the behavior of large groups of individuals, for example by rewarding behavior that is desired. By measuring **feedback** of the manipulation strategy, i.e. how people react, one can find out how "well" the resultant behavior of individuals matches what was desired. Feedback can be anything measurable, e.g. click behavior, purchases, or social media actions. The combination of the fact that data is now everywhere and digital, that measurements of behavior are performed through countless sensors constantly, and that smart algorithms can now deal with huge amounts of data, makes that only *now* tools are becoming available to create what I call *"Walden 3.0."* everywhere.

A key issue in generating the models also becomes important in the exploitation of the models. That is, when models are created that work well *on average*, then the

application of that model to a particular individual may give rise to "wrong" actions. For example, a bank might give out loans based on predictions of trustworthiness of individuals, but as long as optimization is done at the level of the population (i.e. the expected win-loss calculation of the bank for all loans) several untrustworthy individuals will get loans (or vice versa). Supermarkets have excellent opportunities with their loyalty cards for behavioral engineering where manipulation strategies may involve personalized prices. Physical stores are now starting to use WiFi-tracking to identify and follow the behavior of customers to target them with personalized advertisements.

Once measurement (data), prediction, manipulation and feedback are all in place, they open up a possibility for **full experimentation loops** (van Otterlo, 2009) in which data collection, model generation and model exploitation are executed in sequence and indefinitely, enabling algorithms to **experiment** with different settings to see which (kinds of) manipulations work best. Increasingly so, we find ourselves in so-called *digital Skinnerboxes:* information environments set up and controlled beyond our sight, in which we are predicted, measured and controlled. Although most of these experiments remain hidden, the discovery of a recent Facebook case has generated some attention. Kramer et al. (2014) report on how the news feeds of 700.000 people were deliberately manipulated to be either slightly more negative or positive. Results were that negatively influenced people started behaving more negatively on the social network, and vice versa.

D – Information through an Algorithmic Lens

Experimentation loops enable those who have lots of data and computing power to interfere with the daily lives of populations. Recently (van Otterlo 2014c) I have investigated how that influences the access to information and knowledge. Search engines like Google are increasingly assuming the role of **algorithmic gatekeepers** that govern who gets access to what on the web. I developed a metaphor which sees the digital worlds as huge **libraries** governed by a **librarian**, e.g. a search engine. The metaphor shows, for example, how prediction models influence search results, how data of other people biases the information we get to see, and how big data may hinder our direct access to the library, i.e. to information and knowledge. An important thing to note is that large information companies such as Google, Amazon, Facebook, Yahoo, and Twitter are using the increasing amounts of data about us to **bias** the information they provide us with, e.g. in terms of personalized search results, news, news feeds and overviews. That bias has an underlying **business model**, to make money, and is not guaranteed to align with “our” interests.

III – Human Rights and Modern Privacy Concepts

I have briefly surveyed several aspects of the *privacy* concept in the context of big data and smart algorithms. Each of them went beyond the typical conception of privacy as (control over) access to private data, with experimentation loops being the most general and powerful form of surveillance of, and interference with, our lives. This development will only grow stronger because i) increasingly all aspects of our lives become digital, ii) more and more data is measured, collected and generalized with smart algorithms, and iii) we become more predictable and more amenable to manipulation and experimentation every day. Additionally we can see the rise of *platforms* such as Google, Facebook, and Apple that reach out to healthcare, insurance, robotics, education and more, thereby gaining even more data and power.

Human individuals need to be protected against powerful entities. In **the universal declaration of human rights**[1], *privacy* appears in Article 12, making it a fundamental right. However, the formulation of the article resembles the traditional privacy concept of access-to-data and talks about *"interference with his privacy, family, home or correspondence"*. Even though "interference" may be interpreted broadly, this article was written when large-scale experimentation, was unforeseen. The notion of *special data* is interesting in digital context, since, as described, it is very well possible that Facebook data reveals that someone is homosexual. Take a gay individual who is living in a gay-hostile country and has not shared this with anyone. Technically, it would not be a violation of privacy if Facebook uses this information (e.g. for advertisements), since it would be just an *outcome* of a complex statistical model, which *they* generated, with *their* data. In fact, just like the well-known example of a supermarket predicting a pregnancy from purchases, one can even imagine that the gay person is not aware himself/herself of his or her sexual orientation! Because this is not about unauthorised access (since the private data did not exist), why should it be violating human rights?

It seems that Article 12 is not adequate anymore. On the other hand, we immediately touch upon several other articles. The right to not be discriminated is violated on a daily basis since every digital *profile* (e.g. a book-lover, or a muslim who likes comics) discriminates, and determines what we get to see, which offers we get, or how the internet service "sees us". Furthermore, the constant monitoring, sharing and data collection may hinder "intellectual privacy" (see: van Otterlo, 2014c), or *"freedom of thought"* and *"either alone or in community with others and in public or private"* (Article 18). The Facebook emotion experiment exemplifies that freedom of thought can be harmed, even without people knowing. Despite the public outrage about that experiment, it does not seem to be framed as a human rights violation.

"Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers." (Article 19). Being able to exercise this right about political issues becomes difficult once search engines draw you into a *filter bubble* amplifying your initial view. Search results are influenced by statistics of other users, and manipulated for profit, but again it seems hard to frame this as a human rights violation. In addition we should be protected from any form of *propaganda* (war, hate speech, racial, religious) which is hard to ensure, or even detect, when targeted *micropropaganda* hits us every day.

In addition to the Facebook emotion experiment, another experiment was unveiled[2] recently, where millions of Americans were influenced to vote. The researchers admit their experiment may have changed the outcome of the elections. Article 21 is about elections, and while platforms such as Facebook and Google[3] can help campaigns, they can also be utilized to influence and win elections. The Facebook experiments are wake-up calls for supporters of democratic processes, since we do not know which experiments have not been reported and remain undetected. We do know, however, that there are many algorithmic ways to interfere with elections as well.

1 http://www.un.org/en/documents/udhr/
2 http://www.motherjones.com/politics/2014/10/can-voting-facebook-button-improve-voter-turnout
3 Could googling take down a president? http://dl.acm.org/citation.cfm?doid=1327452.1327485

IV – Towards Effective and Practical Digital Human Rights

One of the core ideas in my Walden 3.0 paper (van Otterlo, 2014a) is replacing the in-your-face *Big Brother* style oppression by a softer form of control characterized by small manipulations and nudges, governed by data and algorithms. *Behavioral engineering* in the second form ensures the population is "well-behaved" or "happy", whatever that means, but it enables technology companies to steer society in ways not controlled nor governed by democratic ways. There are new powerhouses in the digital world, and since our lives are becoming increasingly digital, it is time to collectively formulate new rules for these powerful entities, and establish our fundamental rights.

The *Universal Declaration of Human Rights* has a long history[4], ranging from the *Magna Carta* 800 years ago to the now used 1948 declaration. In between several documents have passed that basically established ground rules for powerful entities such as states. Human rights concerns are typically invoked when it comes to states such as North Korea, Iran or Cuba. However, in my view it is essentially about the **power relation** between *any* individual and *any* powerful entity. In digital worlds one may not be able to find pure oppression so easily, but it is simple to point to powerful entities meddling with our lives. And these are the powerful entities that should be bound by new rules.

We should explore the formation of a new declaration of rights, *written for the digital age*. The current declaration was written in a non-digital age, where state powers were dominant. We need a new contract that establishes new rules for new (potential) oppressors and new types of (potential) oppression. I strongly believe that we should move beyond the many partial solutions that come from the limited view on privacy that have focused on *private data*, *data protection*, *encryption*, *transparency, privacy-by-design*, *the right-to-be-forgotten*, *monetizing data* and several legal proposals. I believe that these will not solve the fundamental issues I raise, such as algorithmic versions of gatekeeping, discrimination and experimentation.

I believe there are three fundamental things governments and organizations can do:

- Develop, stimulate and maintain what I call **algorithmic literacy**: in the public and political arena people should *understand* what algorithms are doing, and how they are employed by powerful entities, and for what.
- Explore the contours of **a new declaration of human rights for the digital age**. This includes thinking about fundamental issues of democracy in our neoliberal world, where by default much is left to "the market". However, if private entities are meddling with so many aspects of our lives, aspects that could or should be handled by a democratic process, we should, at least, democratically discuss and decide about them.
- Parallel to the development of a new declaration, we should think about **smart ways for the enforcement of rules**, maybe even through algorithms. Since algorithmic profiling and experimentation is so omnipresent, it may become hard to predict beforehand, hard to detect when it happens, and hard to prove "after the fact". Answering these hard questions about algorithmic interference might show us what is practical, effective and important.

4 http://www.humanrights.com/what-are-human-rights/brief-history/magna-carta.html

Bibliography

Kosinski, M. and Stillwell, D. And Graepel, Y. (2013) Private Traits and Attributes are Predictable from Digital Records of Human Behavior. Proceedings of the National Academy of Sciences (PNAS), 110(15), pp 5802-5805.

Kramer, A.D.I. and Guillory, J.E. and Hancock, J.T. (2014) Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks, Proceedings of the National Academy of Sciences (PNAS), 111(24), pp8788-8790.

Nilsson, N.J. (2010) The Quest for Artificial Intelligence, Cambridge University Press.

Schwartz, H.A. And Eichstaedt, J.C. And Kern, M.L. And Dziurzynski, L. And Ramones, S.M. And Agrawal, M. And Shah. A. And Kosinski, M. And Stillwell, D. And Seligman, M.E.P. And Ungar. L.H. (2013) Personality, Gender and Age in The Language of Social Media: The Open-Vocabulary Approach, PLOS One, Volume 8, Issue 9, pp 1-15.

van Otterlo, M. (2009), The Logic of Adaptive Behavior, IOS Press, Amsterdam.

van Otterlo, M. (2013) A Machine Learning View on Profiling, Chapter 2, pp 41-64, in: Hildebrandt, M. and de Vries, K. (eds.), Privacy, Due process and the Computational Turn – Philosophers of Law meet Philosophers of Technology, Routledge.

van Otterlo, M. (2014a) Automated Experimentation in Walden 3.0. : The Next step in Profiling, Predicting, Control and Surveillance, 12(2), pp255-272.

van Otterlo, M. (2014b) Zo worden we wel heel erg voorspelbaar, Trouw (newspaper, in Dutch), 10th September.

van Otterlo, M. (2014c) The Libraryness of Calculative Devices: Artificially Intelligent Librarians and their Impact on Information Consumption, in:TBA, by Amoore, L. and Piotukh, Routledge, in press.

Wiering M.A. and van Otterlo, M. (2012) Reinforcement learning: State-of-the-art, Springer-Verlag, Berlin, Heidelberg, Germany.