Article

Automated experimentation in Walden 3.0: The next step in profiling, predicting, control and surveillance

Martijn van Otterlo

Radboud University Nijmegen, The Netherlands.
m.vanotterlo@donders.ru.nl

Abstract

Big data technologies are increasingly able to automatically gather data, experiment with action strategies, observe results of such strategies, and learn from their effects. When privacy issues are framed as “control over information” it becomes apparent that some areas in the digital world might be heading to what I call Walden 3.0; communities of interest that are influenced and controlled by measurement and experimentation. Instead of bringing forward Orwell’s 1984 dystopia in the privacy domain as is typically done, I sketch how current developments might be better studied in the context of Skinner’s utopian novel Walden Two. I illustrate several issues through a running example from the domain of artificial intelligence, and by pointing to several areas where automated experimentation can arise. Finally, I raise questions on how to cope with and study the phenomenon of automated experimentation.

Privacy in the Age of Big Data

Given that our society is becoming more digital and more global every day, understanding how privacy issues will change because of these developments is an important matter. Lately the general public is becoming aware, fueled by recent stories about the spying agency, NSA,¹ but also through developments on social networks such as Facebook,² that huge amounts of data about every single person are gathered and stored. The NSA case is typical for the limited view most people still have on privacy: that it is about *access to private data by unauthorized entities*. Many such entities exist as global surveillance systems³ such as Trapwire,⁴ the Domain Awareness System (DAS),⁵ but also FaceBook, Twitter and Google. Still not much attention is drawn to the actual *use* of that information, inducing so-called *data derivatives*

¹ The NSA spying story is by far not new, see the timeline at <https://www.eff.org/nsa-spying/timeline> See also: Want to Evade NSA Spying? Don’t Connect to the Internet, by Bruce Schneier, 7th October 2013, Wired Magazine, <http://www.wired.com/opinion/2013/10/149481/>

² FaceBook tracks micro-actions such as mouse movements as well (see Facebook Tests Software to Track Your Cursor on Screen, by Steve Rosenbush, in CIO Journal, October 30th 2013: <http://blogs.wsj.com/cio/2013/10/30/facebook-considers-vast-increase-in-data-collection/>)

³ Hollywood versions of such surveillance systems can be found in *Person of Interest* (2011) (see <http://www.imdb.com/title/tt1839578/>), *Eagle Eye* (2008) (see <http://www.imdb.com/title/tt1059786/>) and *Enemy of the State* (1998) (see <http://www.imdb.com/title/tt0120660/>). See also Muir (2012) for cinematic connections.

⁴ “TrapWire is a unique, predictive software system designed to detect patterns indicative of terrorist attacks or criminal operations. Utilizing a proprietary, rules-based engine, TrapWire detects, analyzes and alerts on suspicious events as they are collected over periods of time and across multiple locations,” see <http://www.trapwire.com/trapwire.html>

⁵ See at Privacy S.O.S. <http://privacysos.org/2012/august/NYPD-DAS-privacy-ethics>

(Amoore 2011), beyond simple policing or mapping the activities of suspect groups, although it should not be a surprise that the information is gathered for a reason. Quoting Warner and Stone (1970: 124) four decades ago: “Give the administrator in government or business the use of an integrated national population file, with leave to browse through these for selected abstracts of persons in certain income areas, with certain health patterns, education, hobbies, family circumstances—any parameters he wishes—and you provide him with a powerful tool for interference in private lives, to manipulate, to sell more, to condition, to coerce.” In this article we focus on automated algorithms that do just that.

The adagium⁶ of the infamous secret service agency *Stasi* in former East Germany was “to know everything.” Agents covertly gathered information about people, their habits and social circles, and put it in paper files. In contrast, nowadays, the average Facebook user assembles his personal file i) voluntarily and personally, ii) digitally, and iii) connects it to those of people in his social circles. Encouragement for doing so does not only come from companies’ so-called “free” services, but from social pressure. The more control people think they have about privacy, the more eager they are to disclose more information (Brandimarte et al. 2009). In addition, we are constantly tracked by companies and governments; think of smart energy meters, biometric information on passports, number plate tracking, medical record sharing, etc.

In recent years much has been written about the evolution of *privacy* (Tene 2011) in the era of *big data* (Bollier and Firestone 2010; Tene and Polonetsky 2012), *the computational turn* (Hildebrandt and Gutwirth 2008), and *profiling* (van Otterlo 2013); an era called by some *the petabyte age*.⁷ Advances in electronic surveillance,⁸ internet technology and search engines, as well as the rise of social networks, have changed the concept of privacy tremendously. Several recent books elaborate on that, introducing themes and concepts ranging from *Googlization* (Vaidhyanathan 2011), *social networks* (Andrews 2011), *filter bubbles* (Pariser 2011) to *personalization* (Turow 2011). It is not just the amount of data but also novel ways to analyze this data that change the playing field of any single individual in the information battle against big companies and governments. Data is becoming a key element for profit and control, and computers gain in authority. “Individuals are used to second-guessing decisions of bureaucrats and officers but they surprisingly accept without question decisions made by computers” (Zarsky 2012).

A key concept in the privacy literature is *code*: Google’s and Facebook’s services make heavy use of smart *algorithms* developed in artificial intelligence (AI). These learn to predict people’s characteristics and intentions, tailor search results to the individual user’s needs or make search results and recommendations more “social.” Previously (van Otterlo 2013), I described how ideas from automated learning form the algorithmic base of profiling. To understand code in the privacy context one needs to go to the level of *prediction models*, which typically are rich structures combined with statistics. Models are the right level to understand how code can induce (multiple) algorithmic identities (Cheney-Lippold 2011), how they are exploited and how new forms of automated experimentation are starting to influence our daily lives. Google’s biased search results can be mentioned as one example of the latter, but there are many more (van ’t Hof et al. 2012).

I consider privacy to be directly related to the influencing of an individual’s options and perception, or more generally to *control*. Privacy refers to the terms of control over information, not the nature of that information. Novels such as Orwell’s *1984* deal with in-your-face control and archetypical surveillance devices: Telescreens. Contemporary society’s control mechanisms work much more like other books in

⁶ See the motion picture, *Das Leben Der Anderen*, for an interesting interpretation (*The Lives of Others*, <http://www.imdb.com/title/tt0405094/>)

⁷ See the Wired issue: http://www.wired.com/science/discoveries/magazine/16-07/pb_intro

⁸ See the recent stream of news items and newspaper articles on this topic. An example is the text by Naomi Wolfe, “The new totalitarianism of surveillance technology” in *The Guardian*, 5th August, at <http://www.guardian.co.uk/commentisfree/2012/aug/15/new-totalitarianism-surveillance-technology>

the same genre (Claeys 2010): for example, Huxley's *Brave New World*, concerning soft control using entertainment, sex and drugs, but also in terms of Zamyatin's transparent, glass houses, in *We*. However, as I will argue in this paper, technological developments direct us towards a lesser known utopian novel, *Walden Two*, by Skinner (2005 [1948]) which describes a society built on Skinner's ideas⁹ in behavioral engineering.

In this article I introduce the mechanisms that take us far beyond a privacy-as-access point of view towards a privacy-as-control point of view in four steps. Traditional privacy-as-access is about data and control of who gets to see that data, and forms the first step. One step further is about utilizing that data to *infer* additional knowledge: using statistical regularities to predict, with some confidence, various bits of information not contained in the data itself, for example whether someone is female or religious based on his or her Twitter posts. In order to do that, we first need to take a look at how data can be exploited by artificially intelligent algorithms to induce *statistical, predictive models*, which we will do in the next section. However, just predicting new knowledge is not why such predictive models are typically being generated. It is about the *use* of that new knowledge to exploit it for business and surveillance opportunities. In the subsequent section we introduce an extension of predictive models in which it becomes possible to *reason* about *manipulations* of a setting. In this third step the new knowledge obtained from the prediction models is exploited to, for example, optimize advertising decisions in order to influence potential customers and sell many products. The final step then is formed by automating and alternating the generation of prediction models and the manipulations based on those models. This fourth step provides the means to generate Walden 3.0 as I call it: communities of interest which can be controlled through various forms of predictive models and manipulations. The description of this step will be followed by several examples of so-called *digital Skinnerboxes* to illustrate our conceptual model.

Two main contributions are contained within this article. One consists of making explicit the four steps we have just outlined in terms of an algorithmic model from AI. Understanding profiling, control and experimentation at an algorithmic level is vital to understand the evolution of privacy in the digital age. Developing such *algorithmic literacy* may raise more interest in, and awareness of, privacy aspects that go beyond privacy-as-access. The second, most important contribution, is the introduction and development of the Walden 3.0 metaphor to understand how complex, interactive feedback loops of manipulation and behavioral engineering can arise in digital worlds where data and smart algorithms are becoming more and more powerful with each of the four steps we distinguish. Understanding these mechanisms is vital to appreciate the growing power¹⁰ imbalance between powerful and virtually omniscient companies and governments, and an individual.

Profiling, Models and Artificial Intelligence

Models (or profiles) are at the heart of profiling, i.e. the art of constructing predictive theories. This amounts to finding causal rules, invariant patterns and statistical regularities in data. Profiles induce categories, or sub-groups, to which people can belong (to a certain degree); for example, whether they are gay¹¹ or a retirement planner.¹² Recently Kosinski et al. (2013) report on the prediction of various personality traits from only Facebook “likes.” Schwartz et al. (2013) report on similar predictions, now

⁹ The ideas of Skinner have—just like those of Pavlov, Thorndike and Watson—laid a foundation for contemporary algorithms in AI and reinforcement learning (Wiering and van Otterlo 2012).

¹⁰ Note that the main argument of this article ideas are developed starting from recent technological developments. We are aware of a huge literature on “power” and “control” in political, legal and social science discourse. A discussion in that direction is beyond the scope of the current article, but cross connections with that literature should be investigated in the future. An interesting angle for this article can be Foucault's (1979) work: “...power is that some men can more or less entirely determine other men's conduct – but never exhaustively or coercively. A man who is chained up and beaten is subject to force being exerted over him. Not power. [...]”

¹¹ See the research on the MIT reality mining dataset at <http://reality.media.mit.edu>.

¹² These are a subclass of boomer, see advertising at AOL at <http://advertising.aol.com/audiences boomers>

exploiting language use. The very idea that algorithms (or, code) implement such categories and that they can be applied selectively to people, requires a good look at the resulting algorithmic identities (de Vries 2010, 2013; Cheney-Lippold 2011). Algorithms nowadays define how¹³ we are seen, by providing a digital lens, tailored by statistics and other biases.

In this section we will take a look at a small code fragment, just enough to illustrate some notions. However, the underlying ideas are similar to those in huge systems such as Google and Facebook.

Machine Learning of Prediction Models

Artificial intelligence (Nilsson 2010) is the science and engineering of making intelligent machines. *Machine learning* (Flach 2012) as a major subfield is about algorithms that make sense of data. It develops algorithms that can induce (or, learn) theories about a domain from data. For example, based on pictures of apples and pears, a machine learning algorithm learns how to classify (i.e. discriminate between) apples and pears, thereby generalizing this mapping such that previously unseen pictures can be classified too. An active subarea, related to profiling, is *activity recognition* (Yang 2009), often targeting video data, for example to detect suspicious behavior in CCTV data.

An important distinction in machine learning is between discriminative and generative models.¹⁴ *Discriminative* models can only distinguish between individuals. For example, one could construct a model that would classify any given painting into Mondrian and non-Mondrian. It could do this by looking at the painting's colors. Alternatively, *generative* models learn the complete characteristics of the data itself, and are capable of generating specific examples (i.e. Mondrian-like paintings) from this model. This class is more complex, but also more suitable for profiling purposes. For Mondrian paintings, these models would represent aspects such as the average line thickness, the distribution of line lengths and so on. Such models can generate arbitrary predictions about the data. The most prominent class of models are *probabilistic graphical models* (PGM) (Koller and Friedman 2009), which include Bayesian networks.

Another distinction in machine learning concerns the *representation* of data models. Many algorithms see the data as consisting of attributes¹⁵ with their values. For example, a painting can be described as *colorful* = *yes* and *number-of-lines* = 17. A recent trend is to go beyond that and use relational representations (De Raedt, 2008) to allow for explicit connections between data elements, for example,

paintedBy(VictoryBoogieWoogie,Mondrian)

as in a relational database. Such representations are more natural, since most data is fundamentally networked; most particularly when modeling social networks.

Models: Representational Parts

Now, let us start with our example model,¹⁶ in the context of marketing in social networks, represented¹⁷ using the probabilistic relational language *DT-ProbLog* (Van den Broeck et al. 2010). First there are identifiers (denoted in lowercase, e.g. *p1*) of individuals in our data:

¹³ The potential exist for Facebook to even predict breakups between couples based on their activities on the social network. See: Now Facebook Can See Inside Your Heart, Too, by David Talbot, October 27th 2013, in MIT Technology Review <http://www.technologyreview.com/view/520771/now-facebook-can-see-inside-your-heart-too/>
See also: What Facebook Knows, by Tom Simonite on June 13, 2012, in MIT Technology Review, <http://www.technologyreview.com/featuredstory/428150/what-facebook-knows/>

¹⁴ See Andrzejewski et al. (2010) for some graphical examples.

¹⁵ See my previous work for examples of attribute-based models (van Otterlo 2013).

¹⁶ One can find the complete model at <http://dtai.cs.kuleuven.be/problog/tutorial-dtproblog.html>

¹⁷ To make the paper accessible to a variety of readers, we modified notation slightly.

(M1) *person(martijn).*
person(p1).
person(p2).
 ...

In addition, we have information about who is connected in the network to whom (called a trust relation):

(M2) *trusts_directed(martijn,p1).*
trusts_directed(p2,p3).
trusts_directed(p1,p2).
 ...

The data here only represents one-way relations for trust. However, using background knowledge we can define that this relation is symmetric:

(M3) *trusts(X,Y) IF trusts_directed(X,Y) OR IF trusts_directed(Y,X)*

Background knowledge goes beyond the data (i.e. beyond specific individuals) and represents *generic* knowledge: the variables X and Y can stand for any two individuals. A typical example are family relations that, for example, define the concepts parent or grandma (note that one definition uses another here, and generic variables are in uppercase, such as P1):

(M4) *parent(P,C) IF mother(P,C) OR father(P,C)*
grandma(P1,P2) IF parent(P1,P3) AND parent(P3,P2) and female(P1)

Background knowledge also enables representing uncertain (i.e. probabilistic) information. Let us assume we are interested in whether people will buy some unnamed product. In the model we can represent the probability (0.3) that someone will be tempted to buy something if a friend has bought the same product using a rule:

(M5) *0.3 :: buy_trust(Person,Friend).*

This type of probabilistic knowledge is typically learned from data, by (nontrivial) counting in how many cases someone actually buys a product if one of his friends does. This way, statistical information of the data is contained in such rules, but one can also put information explicitly in the model (Thurman and Schifferes 2012); the same representation suffices though. In addition, we also specify the conditions for someone to buy the product:

(M6) *buys(Person) IF trusts(Person,Friend) AND buys(Friend)*
WITH PROBABILITY buy_trust(Person,Friend).

Thus, if some friend buys the product there is a 30 per cent chance (based on the previous rule) that a person will buy the product. Again, note that this is generic knowledge and (without additional rules) applies to all persons appearing in the domain. An interesting aspect is that buys represents one of many *partial identities*: it denotes people that will buy the product because a friend does so. Furthermore, membership of this group of people is probabilistic: there is only a 30 per cent probability that a random friend belongs to that group (unless we know for sure that this person has bought the product). The rules that define them employ the power of *abstraction*: each rule only mentions a few relevant aspects. Note

that “real” identities are not necessary; a generic identifier (e.g. p1) suffices. Identifiers can help¹⁸ in connecting data snippets that are about the same individual.

Summarizing, when we talk about models, we usually mean the generic parts (e.g. M6) that are not tied to specific individuals, i.e. rules denoting under which conditions (types of) individuals have particular properties. Factual information (e.g. M1, M2) about individuals belongs to the data itself.

Models: Algorithmics, Inference and Learning

Data and models are two separate things, although they can be represented similarly. Interaction between them is where algorithms come into play.

Deductive algorithms are about inferring information. From our example model, we can (by logical reasoning) infer that *trusts(p2,p1)* using the generic rule for *trust_directed*. Deduction can be used to infer *novel* knowledge about an individual, for example how likely it is some person will buy the product. Where deductive algorithms focus on logical consequences of a model, *abductive* inference looks for hypotheses and explanations. For example, one could *observe* that someone has bought the product, and finding out what could be—according to the model—the *cause* of that, is an abductive procedure. In our model, part of the explanation for observing that p1 has bought the product can be that he has a trust relation with p2 and p2 has bought the product. *Inductive* algorithms, which include data mining, are central to profiling and represent the means to get from data to models. Induction takes data and generates a model that best fits the data. Some algorithms tune the probabilities for a fixed set of rules based on data, where more general techniques can employ the data to generate the rules themselves. Whereas deduction and abduction are generally called inference techniques and deal with inferring aspects (of individuals), induction is referred to as learning and is targeted at generating or extending a model. In all algorithms probabilistic aspects are needed to deal with the inherent uncertainty in complex domains, but they do make models and computation more challenging. On the other hand, they provide ways to, for example, *rank* hypotheses based on likelihood or express levels of *confidence*.

The Many Faces of Bias

Machine learning is a form of statistical reasoning and therefore all known precautions stay in place (see also Whyte 2004; Reichmann 1961). Aspects such as representiveness of the data, proper data selection, highly improbable events, interpretation of correlations (Tversky and Kahneman 1981), incomplete data and so on, are to be approached with care. Each choice in dealing with data constitutes a *bias* which influences which predictions can be drawn from a model and with what confidence. Note that biases are not necessarily bad; in fact, without¹⁹ bias nothing can be learned because all models would seem equally good. Machine learning bias in the context of profiling has many more aspects (cf. Bozdag 2013; Crawford 2011; Flach 2012; van Otterlo 2013).

In our example model, the most imminent form of bias is *representational*: what to represent, and in which language. For profiling, the choice of which data to track of individuals, which generic relations to consider and how to induce generic patterns from data, bias the generation of models. However, also in the actual *use* of models, considerable bias is present: what to infer (i.e. induce partial algorithmic identities), from whom, when to do that, and, maybe also, why?

¹⁸ Connecting data to the same identifier will become even easier soon when face recognition is employed at a wide scale, both on social network and in camera surveillance systems.

¹⁹ For example, without bias, Google would not be able to function at all: no search engine can index the whole web, and not all hyperlinks are created equally (Vaidhyanathan 2011: 62-63).

Influences on Predictions: Personalization and Socialization

According to Thurman and Schifferes (2012: 11) “predicting accurately [...] at an individual level is a considerable technical challenge.” In AI it is known that very often observable (group) behavior may seem complex, but a prediction model for an individual behavior may very well be simple (Braitenberg 1984).²⁰ This holds especially when these simple-minded individuals interact (Resnick 1994).²¹ The same may hold for other prediction models: it is very likely that simple correlations (as in our example) are enough for the purposes of the model’s maker (e.g. predicting whether someone will buy the product). Accurate predictions about an individual may be difficult indeed, but predicting *on average* in a large population is feasible.

In terms of such predictions, two seemingly opposite trends in profiling can be distinguished: personalization and socialization. The first uses more information about an individual when making predictions, the second uses more information about other people. Both represent a specific form of bias, but from a model’s point of view they are very similar. *Personalization* is using more (demographic, but also behavioral²²) information about a particular individual to tailor predictions to that individual. Examples are Google’s search results based on individual’s cookies or GMail contents,²³ or the use of geographic location (Kim et al. 2011). Personalization happens more and more, and has received recognition through the *filter bubble* principle by Pariser (2011). On the other hand, *socialized search* and collaborative filtering (see also Bozdag 2013) is the use of data of other people to color information. Simple examples are Amazon’s book recommendations, movie suggestions at IMDB, and Google’s social search efforts. The “if you like this then you might like this too”, and the “often bought together” are pieces of information about the book or movie you are currently viewing, based on the models built using opinions and click (and buy) behavior of many other people (see also Alter 2012). Both personalized and socialized models exploit the statistical regularities of data of people *just like you*.

The Exploitation of Data and Models

Our third algorithmic step in this article (following data access, and the building of prediction models) is about the *use* of prediction models, for example to optimize financial profit. Acting upon models is less well studied so far (but see Turow 2011).

²⁰ For example, a robot following a moving light source.

²¹ For example, modeling a realistic school of fish, or a flock of sheep, only three simple rules per individual are needed: i) don’t bump into others, ii) try to match your direction with others, and iii) stay close to the group.

²² Modeling and predicting behavior solely based on externally observable features resembles the behaviorism movement in psychology in the first half of the 20th century. “A much simpler solution [simpler than equating the mind with the brain] is to identify the mind with the [physical] person. Human thought is human behaviour. The history of human thought is what people have said and done” (Skinner 1976). Behaviorists felt that because it was absolutely impossible to look inside someone’s brain to find out why some behavior comes about, they argued that psychology should be only concerned with externally observable features of behavior. Stipulating higher-order motivations, reasons, goals, intentions and things like that are merely speculation about the reasons for certain behavior. Whereas humans could possibly talk about their motivations, and perform introspection on what was on their mind (as cognitive psychologists would do) is not the right way to do repeatable, falsifiable science, especially with animals which are not capable of verbal explanations or introspection. The success of search engines such as Google and many other applications in activity recognition show that much can be done by “just” correlation patterns. The behaviorist movement in psychology is mirrored in AI by movements such as behavior-based approaches and new AI, and also reinforcement learning (Wiering and van Otterlo 2012). These directions often place a particular emphasis on not representing, or reasoning about, intentions, goals, motivations, or inner states, but instead focus on how predictions and behaviors directly follow from observable behavior data. This is unlike other trends in AI where the goal is—in contrary—to model cognitive belief states explicitly. It is good to keep in mind that many of the current techniques in data science which we describe here can in principle be applied in both contexts, although we lean slightly towards the behaviorist approaches here. For more historical developments in artificial intelligence see the excellent book by Nilsson (2010).

²³ See <https://www.google.com/experimental/gmailfieldtrial>

The Use of Data and Models

Models and profiles are potentially influential when they are invoked. Much has been written about possible dangers for privacy when it comes to inferring new knowledge from profiles. As Anrig et al. (2008) put it: “Profiles discovered by these techniques may not be anticipated, desired or politically correct but modern algorithms do not (yet) have the power to care.” Imagine a social network containing many individuals who have bought the product. Now, the chance that I am not influenced by any one of them (according to the model) becomes smaller as the number of people in my network increases. Due to this phenomenon, I could be placed in the very-tempted-to-buy category, without me knowing that, and maybe treated differently based on this categorization. I seem to have no control over how I am perceived (i.e. what my algorithmic identity is according to a particular model) since I do not know that there is such a model, what it looks like, or when it is invoked (i.e. I am being treated differently). The influence an individual has about how and when profiles are being used is very limited. The same holds for control over your own data. Since models are built from data of many individuals, and your data is as good as your neighbors’, often very few data items about you are needed to infer new knowledge. This renders many endeavors directed at protection, anonymization and even revocation of data relatively useless. All relevant (statistical) knowledge about individuals is already included in the models. One could say that the persons behind the data are forgotten through the model.

The Value of Inference in Models

Models enable attaching a probability to certain attributes, facts or complete situations. A first step to modifying this probability distribution to our advantage is by explicitly modeling certain factors we can influence. In our example, we assume that we can send an individual targeted marketing in order to influence that person’s likelihood of buying the product. Let us assume that we learn from observed data that targeted marketing induces a 20 per cent chance that a person buys the product, inducing the additional rule:

(M7) *0.2 :: buy_marketing(Person).*

Now, this creates additional possibilities to express (in the model) when somebody will buy the product. In this case that is the following rule saying that a person will buy the product (with probability 0.2) if he gets a targeted advertisement.

(M8) *buys(Person) IF marketed(Person)
WITH PROBABILITY buy_marketing(Person).*

Together with our original model now two factors may influence someone’s buying behavior, and we assume one of them is under our control:

(M9) *IF person(Person) THEN marketed(Person) IS yes OR no.*

This rule denotes whether we send advertisements to whoever is filled in for the variable Person. Note that this single rule represents a possibly huge number of decisions; one for each individual. Now, having in place the means to express statistical information and a way to influence the behavior of the population through decisions, we need additional elements to express aspects of “good” and “bad” decisions through a reward function²⁴ (or, cost function). It specifies for any particular state of the world how much (parts of) that state is “worth.” In our example model, it is natural to specify that every time a person buys our product, we earn money, i.e. reward in this model is directly related to monetary value. On the other hand, targeted advertisements will cost money, and are represented by a negative reward.

²⁴ See for a more elaborate discussion (van Otterlo 2009).

(M10) *IF person(P) AND buys(P) THEN REWARD IS 5.*
IF person(P) AND marketed(P) THEN REWARD = -3.

Here it is natural to express costs and rewards in terms of money. However, in any domain some kind of costs may be specified. For example, in security applications, the occurrence of a bomb explosion is reflected as a huge negative cost (if it would occur), and the strip search of an innocent civilian should (at least) count as some negative cost. In this way, the trade-off between security and inconveniences is explicitly stated in the model. In yet another domain one could trade-off the number of votes against the amount of flyers printed in an election scenario. In this way, data about individuals can be monetized.²⁵ Note that rewards and costs here are at the level of the *global* model. I need to stress here that rewards at the (local) level of individuals are only implicitly in the model in the form of *incentives* coming from the actions. For example, supermarket discounts (an action, but also a cost) could persuade (local reward) an individual, who then possibly buys the product (a global reward). Extending²⁶ models to explicitly deal with individual rewards is possible but outside the scope of this article.

Global Optimization: local consequences

Our new model explicitly represents negative and positive effects (rewards). However, because of the probabilistic aspects of the model, one can never say that a certain situation—in our case which specific people buy the product—*will* happen. What we can quantify is what the probability is with which a certain situation can happen, and in addition, what would be the costs and profits for that situation. Thus we need to talk about *expected*²⁷ rewards: utility. It is defined in terms of i) the total profits of a given situation minus the total costs of that situation, *multiplied by* ii) the probability that this situation might occur. Taking the sum of all possible situations—no matter how unlikely—will produce the expected value, or the utility. In our example domain, we can compute the expected value of a policy, i.e. a strategy for marketing a subset of the population. Let us assume that our strategy is:

(M11) *[marketed(martijn), marketed(p4), marketed(p27), ...]*

The decisions, i.e. who gets a targeted advertisement, cause one of many possible situations to occur. For each of them, we can compute the probability. For example, in some of them the person p4 will influence some other person, say p13, with probability 0.3 and in other situations he will not. Overall, the strategy's expected utility expresses the outcome of the trade-off between costs for targeted advertisements and the expected amount of products sold.

Finding the globally optimal strategy, i.e. that strategy with the highest expected utility, is far from trivial given the many possible strategies. In essence it is abduction: find a possible cause (i.e. a marketing strategy) for a situation in which much profit can be expected. There are many types of bias at work: the best strategy is dependent on the population in the social network, the structure of that network, and all (probabilistic) rules in the model. In fact, we should emphasize that the most important bias in this type of systems is the reward function²⁸ (see also van Otterlo 2009; Wiering and van Otterlo 2012). By valuating things differently one gets different solutions of what to do.

²⁵ Recent work by Spiekerman on “privacy as property” shows how people are forced to value their privacy exactly (http://www.wu.ac.at/ec/faculty/spiek_pres_propertytielburg2012). If we extend the utopian literature connection in this paper even further we might end up in Ayn Rand's novels where literally all aspects of life are monetized.

²⁶ For example, the field of *multi-agent reinforcement learning* studies both the global learning behavior as well as individual learning performance of rational agents in a group (see more on this in chapters 14 and 15 of Wiering and van Otterlo 2012).

²⁷ The expected number of eyes on a die is $(1+2+3+4+5+6)/6 = 3.5$ but you will never actually throw this number.

²⁸ This is equally important in computational reinforcement learning. For example; to teach a robot how to escape from a maze, we give it a small negative reward for every step and a big reward if it gets out. These *incentives* nudge the robot into learning to go to the exit as quickly as possible.

From a privacy-as-control point of view an issue arises with the globality of models and the locality of an individual. The fact that models are optimized at a global scale may very well make the consequences for an individual incomprehensible, meaningless, or even disruptive. In our example, the globally optimal strategy is a selection of people who get marketed. Now, depending on the social network relations, it might be optimal to target person p47 too because in the complicated probabilistic interactions that follow from applying the optimal strategy the most (expected) profit will be made. However, from that individual's point of view, it might not make much sense since this person might not have the intention at all to buy the product or to influence his friends in doing so. Whereas advertisements might be ignored, the very fact that decisions are made at a global scale that have consequences for what individuals observe or can do at a local scale is another example of the power imbalance of knowledge and privacy.

Towards Walden 3.0—Automated Profiling and Experimentation

The tools described in the previous sections provide the building blocks for powerful surveillance machines. The first tool is *model generation*, which takes observed data about a population and builds a general model (such as in M5-M8). The second tool is *model use*, which takes a model (and an accompanying reward function as in M10) and computes the optimal way to exploit the information in the model to act (i.e. M9 and M11) upon individuals, for example by targeting them with advertisements²⁹ or by re-ranking their search results. Now, conceptually it is good to separate these tools, but in practice there is a constant dynamic of new data coming in, revised models being built, reward functions being adjusted, and actions being performed based on the new models. One could say that model generation takes data and delivers a model, whereas model use takes a model and causes new data to arise. The resulting *profiling loop* can be applied to any domain where data is digital and ubiquitous. Interestingly, this *feedback loop* can be driven using human intervention, but can also function in an automated fashion driven purely by algorithms. Google and Facebook can do this with populations larger than most countries, and they typically do (Olsthoorn 2012). This *experimentation* loop represents our fourth and final algorithmic step in this article.

It is here where Skinner's Walden Two fits better than the traditional Big Brother. Walden Two depicts a small society in which principles of behaviorism are applied, and where the population is conditioned by manipulating how individuals get rewarded (or, to a lesser extent, punished). Good behavior and happy citizens are obtained not by force, but by manipulation. Automated, real-time profiling can form the basis of a very ambient, very covert form of control using lots of data and smart algorithms. Quoting Skinner (2005 [1948]: 192–193):

No, the potency of behavioral engineering can scarcely be overestimated. It makes one wonder why the techniques haven't been put to better use long before this. We could teach our children to be satisfied with a very limited and rigorous existence, to despise other forms of society, and to turn from the pleasures of the flesh. We might make such a society last for many years.

Various forms of this behavioral engineering can be implemented in a digital world. For example, Google conditions us to accept and believe that a search results list does in fact deliver what we want, and Facebook users are nudged into filling their timeline, even about the period before they actually had an account. Global actions provide the local incentives individuals use to change their behavior. It all depends on how the reward function of the model is set up and how individuals react to changes in their personal information environments. Again quoting Skinner: "We aren't satisfied to produce merely a happy people. Our technology is powerful enough to make men happy under many conditions of life." Walden Two was published a long time before any realistic form of digital surveillance was possible.

²⁹ See Greengard (2012) who covers the typical case of pregnancy prediction based on purchases.

Based on the technology available, the emergence of a “Walden 3.0” with control using positive reinforcements and behavioral engineering seems a natural development.

Automated Learning and Manipulation: Experimentation

Once model generation and model use are in place, all is set for automated *experiments* with profiled subjects. For example, a webshop may vary its prices to see how much each (type of) individual is willing to pay for certain products. A politician’s campaign machine may experiment with various interaction methods (advertisements, personal approach, etc.) to see what works best in order to get the most votes. A web search engine may experiment with the returned search results, and their order, to ensure many users will follow sponsored links.

The general structure for automated experimentation is the following. Based on the behavior of individuals in some population, profiling models are generated. Then, combining these models with a reward function on which behaviors of individuals are desired (e.g. profitable) a globally optimized strategy, i.e. the one with the highest utility, can be computed according to which the individuals can be targeted, manipulated or influenced, for example through sales offers or by providing specialized information. This strategy will then trigger changes in the behaviors of individuals, which can be tracked and measured using the same methods as before. The crucial point here is that because so much of individuals’ behavior is uncertain (i.e. probabilistic) optimized strategies will have to be evaluated using an *expected* profit. But because there are many individuals in most systems, it pays to *experiment* with strategies to see which one works best. This might cost more than it delivers in the short run, but may give rise to much better strategies in the long run. In terms of computational learning theories (derived from human and animal learning studies) this amounts to the *explore-exploit tradeoff* (Wiering and van Otterlo 2012: Chapter 1): one needs to *exploit* profitable knowledge, but one needs to *explore* to find potentially better strategies.

Even though many services on the internet can be studied in terms of automated profiling (or adaptive hypermedia systems, cf. Steichen et al. 2012), most are too large to study and inaccessible due to models and algorithms which are kept secret.³⁰ One fully automated system that is still compact enough to understand fully is the *robot scientist* (King et al. 2004, 2009). It uses a language similar to our running example (Van den Broeck et al. 2010), and it automatically performs scientific experiments concerning the growth of yeast. The system has a current model, does abduction to come up with interesting experiments, and can perform actual experiments on yeast samples. Based on observed behavior the system can learn additional rules in the model and continue with new experiments, incorporating the costs of experiments (i.e. a reward function). A mental step from yeast experimentation to data-driven social science is not too hard to make.

Another insightful example of a general form of experimentation comes from *Google’s A/B testing* facility (Christian 2012) in Google Analytics.³¹ It enables the testing³² of possibly hundreds of different versions of a webpage on a huge audience to see how each triggers different reactions of individuals. For example, a store could experiment with which color website they sell the most products, and learn from customer’s reactions, aided by profiles built from their purchases and click behavior. People are usually unaware of their participation in such ambient internet Skinnerboxes.³³

³⁰ Interestingly, the Dutch election system has gone back to the red pencil just because the algorithmics of voting machines was not transparent <http://wijvertrouwenstemcomputersniet.nl>

³¹ See <http://www.google.com/websiteoptimizer/tutorials.html>

³² See also the *website morphing* technique described by Hauser et al. (2009) in which websites are adapted to the (inferred) cognitive style of the user, in real time.

³³ Skinner was famous for his experiments in a so-called Skinnerbox, a form of puzzle box in which animals were taught some specific skill, see http://en.wikipedia.org/wiki/Operant_conditioning_chamber

Recent advances in experimentation come from *probabilistic programming languages* (Poole 2010). Unlike traditional programming languages, these support learning and probabilistic inference as built-in capabilities. For example, one could write down parts of a model and action strategy in a program (e.g. our running example) and probabilities (and rewards) are estimated automatically from data. A programmer can write down a *partial program* after which automatically missing parts are filled in using data. Interestingly they could even learn from trial-and-error; i.e. the program could experiment with some strategies and based on the feedback it gets of “how well” things are going according to the reward function definition, it may alter its behavior. Computationally this kind of learning is known as *reinforcement learning* (van Otterlo 2009; Wiering and van Otterlo 2012). Ideas in this area came from, among others, Skinner himself, and are now used for many intelligent learning tasks, for example a robot making pancakes (Lockerd-Thomaz and Breazeal 2008).

Skinnerboxes: Domains for Automated Manipulation

In this section we briefly point to several domains where profiling and experimentation happen.

Viral marketing, Advertisements and Shopping

Our running example deals with viral marketing. Goldfarb and Tucker (2011) argue that targeting is very effective: “...our estimates suggest that one could reasonably expect a large drop in advertising effectiveness for consumers who choose to opt out of targeting.” From a modeler’s point of view, the goal (and reward function) is clear, lots of data comes in from buying behavior, and models can be tested quite objectively. For example, Fong (2012) reports on how targeted marketing can influence search activity and willingness to try alternative products. Amazon and other webstores are probably the most visible occasions where we really notice that our data is being used, through recommendations and advertisements. The same goes for loyalty card systems in, for example, supermarkets. One could say that data has become so important that it has changed our view on consumer behavior tremendously. (Zwick and Knott 2009: 240): “We argue that the constant and compounding growth in the volume of data coupled with the rising analytical powers of computers has endowed the customer database with an immediate strategic importance in a company’s economic value creation process.” Interestingly, Palmås (2011) talks about the same kind of issues we discussed earlier when it comes to global optimization and local consequences of models:

[t]his allows companies to control markets in new ways: instead of being generous to loyal customers, they can shift their attention to the potentially disloyal customers. Data mining not only enables them to single out customers who are statistically profitable, it helps them calculate the exact minimum level of getting consumers to stay loyal.

Finding the global optimal policy in our running example exactly amounts to that: global actions (M9) can provide local incentives (i.e. local rewards) that may, in turn, generate global rewards if someone buys products (M10). Actions cost money, for example because they lower prices for a group of people, or they involve costly marketing (M10). The prediction model (M5-M8) provides the means to predict whether someone is likely to buy something, under which circumstances. Experimentation in shopping domains is a natural step and already happening.³⁴ Supermarkets know³⁵ their (individual) customers very well and know which local incentives have a high chance of nudging someone into buying. However, nothing is sure: what they can do is do controlled experiments in which they perform some actions (e.g. lower prices), estimate updated models on the behavior of people, and see which actions generated the most profit. Prediction models such as M8 might need to be refined: maybe a subclass of these persons shows a

³⁴ Since Albert Heijn knows so much about their customers and is able to maximize profits using that knowledge, obviously complaints have risen about discounts not being discounts at all, see (in Dutch)

<http://www.distrifood.nl/Formules/Algemeen/2013/10/AH-houdt-34-miljoen-keer-bonusvoordeel-in-eigen-zak-1394224W/>

³⁵ See: Tesco using Minority Report-style face tracking technology so ads on screens can be tailored, by Paul Cockerton, 4th November 2013, Daily Mirror, <http://www.mirror.co.uk/news/uk-news/tesco-using-minority-report-style-face-2674367>

higher likelihood to buy products if offers are being made around pay day. Such analyses may also provide insights into how to balance the rewards-costs trade-offs (such as in M10) if much more complex models, with various components, need to be exploited.

Search Engines

As Vaidhyanathan (2011: 84) says: “Google is a system of almost universal surveillance, yet it operates so quietly that at times it’s hard to discern.” Search engines, in particular Google, are becoming powerful profiling machines. They typically implement all profiling techniques discussed in this article. Some are visible, such as Google’s search query completion, whereas others are much more hidden, such as the influence of Gmail or Google+ circles on search results. An important aspect of search engines is that they function as *gatekeepers* (Introna and Nissenbaum 2000; Granka 2010) and can influence and color³⁶ the information individuals get about the world. Goldman (2011), Mager (2012) and Bozdag (2013) describe the many forms of bias in search engines. Mager explicitly sees dominant capitalist values becoming embedded in the search results. Goldman (2011) addresses the issue of portalization: where Google previously wanted to refer users to third-party websites, it increasingly wants to do more, such as giving aggregated information, or it wants to provide more, such as mail, videos, images etc. The more it can assume a portal function, the more data it has for building prediction models. Furthermore, “Google is moving from simple data retrieval to a system that understands how we think and what we want—before we even know we want it” (Vanderbilt 2013); a clear hint towards automated bias generation and manipulation. The mentioned A/B testing facility is a role model of the experimentation described in this article. Global actions can be anything dealing with the generation of (the layout of) a webpage, where the connection with local incentives may come from the field of design e.g. which colors and fonts do people prefer. The cost and reward model come from revenue models; most search engines thrive on click behavior (local actions and a predictive model of it like M8). It is interesting to realize that every website we visit may be an experiment, explicitly designed for us (or people like us).

Surveillance on the Workfloor

A company can be seen as a small-sized society, and increasingly workers are being monitored through the use of surveillance cameras, tracking devices (in logistics), or e-mail. Recently, Microsoft applied for a patent³⁷ about what some call “human resource software from hell.”³⁸ The basic idea (from the first paragraph of the patent text) is to create “[a] method to be executed at least in part in a computing device for monitoring, analyzing, and influencing organizational behaviour.” The main goal of the system is “... providing analysis results such that desired behaviors are encouraged and undesired behaviors are discouraged.” How much more “Walden” can it get? Companies lend themselves very well for profiling systems because more control can be obtained in a limited setting where individuals are more dependent on the profiler and consequences of “bad” behavior can be more severe. Predictive models (M8) can be built on the basis of (e-mail) communication, working hours, work done (when working in the company cloud), and much more. The global cost model (M10) and actions (M9) are to be decided by what the company wants them to be. The patent is not very specific³⁹ about it, but no doubt disciplinary

³⁶ Recently, Google has started censoring so-called torrent links in its search results. This is being done under pressure concerning copyright infringements, but this may affect the visibility of other, legitimate content, thereby limiting access to information. See <http://www.zdnet.com/au/google-takes-small-step-against-online-piracy-7000002490/>

³⁷ The author was interviewed by Jolein de Rooij about this patent. This interview was published in Intermediar-PW, January 2012 (<http://www.intermediarpw.nl>).

³⁸ http://www.theregister.co.uk/2011/11/18/microsoft_patent_employee_monitoring/

³⁹ The Dutch city of Eindhoven is starting a project to monitor and manipulate a popular recreation area in the city center, but here too the goals to be reached are relatively unclear or unspecified. See the “Living Lab” project (in Dutch), <http://www.bd.nl/regio/brabant/eindhovens-stratumseind-wordt-proeftuin-1.4061962> Similar projects exist, for example in the UK. It is possible to use trash cans to track people's phones to get an accurate measure of what is happening in an area. See: No, this isn't a scene from Minority Report. This trash can is stalking you, By Dan Goodin, Ars Technica, 9th August 2013, <http://arstechnica.com/security/2013/08/no-this-isnt-a-scene-from-minority-report-this-trash-can-is-stalking-you/>

measurements are among them. Experimentation could be done by enforcing several policies, for several time periods, and testing which resulting personnel behavior fits⁴⁰ the company's expectations.

Reading, News, Education and others

The way in which news and information about the world in a digital world can be biased by personal preferences or by social structures is unprecedented. Messing and Westwood (2012) describe how social media change what we read and why. Thurman and Schifferes (2012) discusses the increased personalization of news and relates it to Pariser's filter bubble. The rise of *social* digital reading practices is a recent phenomenon (Richards 2013; Alter 2012). Overall, the consequences of automated profiling on what people read about the world could have tremendous consequences for society. Another area that employs profiling machines is elections, where global reward functions optimize vote counts. The influence (i.e. providing local incentives) of data, polls, social networks and search engines could be significant (Conti 2008; Rustin-Paschal 2011). Evans (2012: 884-885) says that political data-mining can be done nowadays with relatively little effort and cost; a data-miner can compile a useful voter database from just publicly available data. A much more recent domain where large-scale experimentation is to be expected concerns the *massive open online courses* (MOOC) that are now being developed by many universities (Carr 2012). Profiling of students, through measuring of their performance and effort in these course, and their use of the system, induces profitable data with which companies who start offering such systems will create value of. Moving to the physical world, it is expected that robots and drones will have a huge impact on data collection, privacy and control in the near future (Sharkey 2008). These are just some additional examples next to the game industry, the entertainment industry, cloud computing, and many more.

Discussion, Conclusions and Outlook

We have seen four steps of increasingly more intelligent algorithms influencing how we are being treated, categorized and manipulated in digital domains. Our final model brought us to Walden 3.0, in which automated experimentation can be performed based on an underlying global business model that can be very different from an individual's local view. The interplay between data and models, and between representation, bias and algorithms, are central to the rise of prediction machines founded on the algorithmic basis of machine learning. What happens at the local, individual level in terms of manipulation is incomprehensible without access to how at a global level information is exploited, and according to which reward function. Now the question is how good or bad it is that experimentation and manipulation happen in the digital world. In other words, is Walden 3.0 utopian or dystopian, or both? After all, there are huge benefits of personalization, social search and global optimization of search engines. Quoting Vaidhyanathan: "Google never promised to be comfortable and benign: it just promised not to be evil, whatever that means" (2011: 75). In our view of privacy as control over information a growing power imbalance exists between an individual and prediction machines. The fact that your individual life is determined by how algorithms categorize you and how they act upon that makes you want to know (at least) when that happens, how that happens, and why.

Previous works have discussed technical solutions against privacy loss such as privacy-preserving techniques, privacy-by-design, or various forms of obfuscation (Brunton and Nissenbaum 2011). Whereas that may sometimes work, they all are vulnerable and do not solve all problems, since data can increasingly be linked. For example, anonymous data from smart energy meters⁴¹ might get de-anonymized for people with public agendas (e.g. scientists). Other examples include the de-anonymization

⁴⁰ The 2013 BBC documentary *The secret life of the cat* ends with a discussion on how domestication has changed the behavior of cats in our society. The project tracks a small, localized society of roughly 50 cats in their daily routine using cameras, GPS, etc. <http://www.bbc.co.uk/mediacentre/latestnews/2013/secret-life-of-the-cat.html>

⁴¹ See this white paper on some other options for privacy and smart meters by George Danezis http://research.microsoft.com/en-us/projects/privacy_in_metering/privacytechnologyoptionsforsmartmetering.pdf

of AOL data (Barbaro and Zeller Jr. 2006) and statistical prediction of social security numbers (Acquisti and Gross 2009). Among existing remedies (Schermer 2011; Tocha et al. 2012), demanding *transparency* (Hildebrandt and Koops 2010) seems natural; we would like to know how and when we are profiled. Two problems prevent transparency from being a solution. The first is that governments and companies own the models and are not eager to share information because of security or commercial reasons. Models are based on statistics they gathered, and no (legal) mechanisms are in place to force them to share that information. The second problem is more severe: let us assume we do get all information related to your search results or your news items for today, how would we convey this information? Any reasonably sized model could produce many hard-to-understand explanations. In addition, the probabilistic information should be accompanied by the underlying statistical data, biases, settings of specific machine learning algorithms and so on. Such information would be impossible to understand for humans.

Thus, even when considering less (i.e. than profiling and experimentation) severe privacy threats, it seems that standard solutions do not suffice. Maybe it points to the fact that if *technical* countermeasures are appropriate, they should be just as smart and especially adaptive as the profiling and experimentation engines themselves. Artificially intelligent algorithms that automatically behave in such a way that they help the user in the control of information (i.e. privacy) and that can assist the individual in digital Skinnerboxes are an interesting direction which should be explored in follow-up research. If Google can invent automated chatbots⁴² that autonomously engage in inter-personal interactions on social networks, we can think of other types of bots that automatically represent us in digital domains, possibly negotiate about trading an individual's data, see through possible manipulations of other bots (e.g. those of Google or Facebook or the government), or otherwise assist us in our digital society. Big tech companies are investing⁴³ heavily in smart technology and AI, and it only seems natural that users of those systems should get help from the same technology (but with a different business model). However, technology as a solution to problems caused by similar technology seems like a good idea, but, for example, in the automated trading⁴⁴ area, this has resulted in hard-to-understand and hard-to-control systems.

Solutions for privacy-related problems must also come from political action. Our tacit agreement to allow big tech companies and governments to digitally profile us and manipulate us using business models not known or comprehensible at the level of the individual, has not been given much thought. Democratic debate and decision are needed to make informed decisions about the kind of technological society we want. Is it ok for supermarkets to have the power to manipulate individuals far beyond the level of traditional advertisements? Is it alright for rich political parties to be able to influence individuals because they have better technology instead of better ideas? Is it ok for search engines to experiment with the news item they provide you with just for the sake of optimizing their global business model? Should we accept automated bots on internet forums that can influence or steer important discussions? What systemic risks arise if such control is enforced over many individuals? What happens to society if huge amounts of customers, voters, students or general, informed citizens can be predictively biased by technology capable of profiling and experimentation on a wide scale? These are relevant questions and they deal with the issue of *control*: the more powerful your technology is, the more control you have. More specifically, what is important is the gradual shift in who is in control. Kuhlmann, speaking about Skinner, stated:

In his view, wanting to choose freedom from control means to cling with “nostalgia” to an outdated concept of human nature—which at this point in history is not only naive but also dangerous because it stands in the way of redesigning western culture with a more

⁴² <http://www.dailymail.co.uk/sciencetech/article-2515635/Google-files-patent-robot-writes-Facebook-posts-emails-tweets--need-FULL-access-scan-accounts.html>

⁴³ E.g. See <http://www.wired.com/wiredenterprise/2013/12/facebook-yann-lecun-qa/>

⁴⁴ Related phenomena are automated trading systems where most human intervention is removed, mainly due to the speed of acting upon the data. See a story about hypertrading and what can go wrong at: <http://www.bbc.co.uk/news/magazine-19214294>

rational and scientific understanding of human nature in mind. “To refuse to accept control,” Skinner argued repeatedly, means “merely to leave control in other hands.”...
(2005: 28)

Political and democratic decisions should be at the basis of such a shift of control in other hands. The abovementioned recent NSA case makes the general public aware of some forms of growing power imbalances but a much broader discussion is needed on the type of technological developments we have described in this article: profiling and experimentation. This includes questions like: how can we detect (and cope with) the issue of experimentation (e.g. discovering living inside the Matrix⁴⁵) and how does it relate to free will and autonomy? This article aims to contribute to algorithmic literacy and becoming aware. “You’re getting paranoid, Carter. That’s a step in the right direction.”⁴⁶

References

- Acquisti, A. and Gross, R. 2009. Predicting Social Security Numbers from Public Data. *Proceedings of the National Academy of Science* 106: 10975–10980.
- Alter, A. 2012. Your E-Book is Reading You. *The Wall Street Journal*, July 19th. Available at: <http://online.wsj.com/news/articles/SB10001424052702304870304577490950051438304> (accessed May 7, 2014).
- Amoore, L. 2011. Data Derivatives: On the Emergence of a Security Risk Calculus for Our Times. *Theory Culture Society* 28(6): 24–43.
- Andrews, L. 2011. *I Know Who You Are and I Saw What You Did*. New York: Free Press.
- Andrzejewski, D., Stork, D. G., Zhu, X. and Spronk, R. 2010. Inferring Compositional Style in the Neo-Plastic Paintings of Piet Mondrian by Machine Learning. *Proceedings of Electronic Imaging: Computer Image Analysis in the Study of Art (SPIE)* 7531, Computer Vision and Image Analysis of Art, 75310G. Available at: <http://spie.org/Publications/Proceedings/Paper/10.1117/12.840558> (accessed May 7, 2014).
- Anrig, B., Browne, W. and Gasson, M. 2008. The Role of Algorithms in Profiling. In: *Profiling the European Citizen: Cross-Disciplinary Perspectives*, eds Hildebrandt, M. and Gutwirth, S., 65–87. Netherlands: Springer.
- Barbaro, M. and Zeller Jr., T. 2006. A Face is Exposed for AOL Searcher No. 4417749. *New York Times*, August 9. Available at: <http://www.nytimes.com/2006/08/09/technology/09aol.html> (accessed May 7, 2014).
- Bollier, D. and Firestone, C. M. 2010. *The Promise and Peril of Big Data*. Washington, DC: The Aspen Institute.
- Bozdag, E. 2013. Bias in Algorithmic Filtering and Personalization. *Ethics and Information Technology* 15(3): 209–227.
- Braitenberg, V. 1984. *Vehicles: Experiments in Synthetic Psychology*. Cambridge, MA: MIT Press.
- Brandimarte, L., Acquisti, A. and Loewenstein, G. 2009. Privacy Concerns and Information Disclosure: An Illusion of Control Hypothesis. In: *Proceedings of the 9th Workshop on the Economics of Information Security (WEIS 2010)*, June 2010.
- Brunton, F. and Nissenbaum, H. 2011. Vernacular Resistance to Data Collection and Analysis: A Political Theory of Obfuscation. *FirstMonday* 16(5).
- Carr, N. 2012. The Crisis in Higher Education. *MIT Technology Review* 115(6): 32–40.
- Cheney-Lippold, J. 2011. A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control. *Theory, Culture & Society* 28: 164–181.
- Christian, B. (2012) The A/B Test: Inside the Technology that’s Changing the Rules of Business. *Wired Magazine*, 25th April. Available at http://www.wired.com/business/2012/04/ff_abtesting/
- Claeys, G. ed. 2010. *The Cambridge Companion to Utopian Literature*. Cambridge: Cambridge University Press.
- Conti, G. 2008. Could Googling take down a president? *Communications of the ACM* (1): 71–73.
- Crawford, K. 2011. Six Provocations for Big Data, paper presented at Oxford Internet Institute’s – A Decade in Internet Time – Symposium on the Dynamics of the Internet and Society (available at <http://ssrn.com/abstract=1926431>).
- De Raedt, L. 2008. *Logical and Relational Learning*. Berlin and Heidelberg: Springer Verlag.
- de Vries, K. 2010. Identity, profiling algorithms and a world of ambient intelligence. *Ethics and Information Technology* 12(1): 71–85.
- de Vries, K. 2013. Privacy, due Process and the Computational Turn. In: *Privacy, Due process and the Computational Turn – Philosophers of Law meet Philosophers of Technology*, eds Hildebrandt, M. and de Vries, K., 9–38. London: Routledge.
- Evans, C. 2012. It’s the Autonomy, Stupid: Political Data-Mining and Voter Privacy in the Information Age. *Minnesota Journal of Law, Science and Technology* 867: 867–898.
- Flach, P. 2012. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. Cambridge: Cambridge University Press.

⁴⁵ Interesting movies about the reality we live in and the effects of living inside a simulation are *The Matrix* (1999), *The Thirteenth Floor* (1999) and *eXistenZ* (1999).

⁴⁶ Quote from the television series, *Person of Interest*, (<http://www.imdb.com/title/tt1839578/>) season 1, episode 12. Person of interest is about a huge prediction machine for surveillance.

- Fong, N. M. 2012. Targeted Marketing and Customer Search, working paper series. Temple University Fox School of Business; MIT Sloan School of Management. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2097495.
- Foucault, M. 1979. Omnes et Singulatim: Towards a Criticism of Political Reason. In: *The Tanner Lectures on Human Values* (October 1979).
- Goldfarb, A. and Tucker, C.E. 2011. Online Advertising, Behavioral Targeting, and Privacy. *Communications of the ACM* 54(5): 25–27.
- Goldman, E. 2011. Revisiting Search Engine Bias. *William Mitchell Law Review* 38(1): 96–110.
- Granka, L. A. 2010. The Politics of Search: A Decade Retrospective. *The Information Society* 26(5): 364–374.
- Greengard, S. 2012. Advertising gets Personal. *Communications of the ACM* 55(8): 18–20.
- Hauser, J.R., Urban, G.L., Liberali, G. and M. Braun. 2009. Website Morphing. *Marketing Science* 28(2): 202–223.
- Hildebrandt, M. and Gutwirth, S. eds. 2008. *Profiling the European Citizen: Cross-Disciplinary Perspectives*. The Netherlands: Springer.
- Hildebrandt, M. and Koops, B.J. 2010. The Challenges of Ambient Law and Legal Protection in the Profiling Era. *The Modern Law Review* 73: 428–460.
- Introna, L. D. and Nissenbaum, H. 2000. Shaping The Web: Why the Politics of Search Engines Matters. *The Information Society* 15: 169–185.
- Kim, B., Ha, J.-Y., Lee, S., Kang, S., Lee, Y., Rhee, Y., Nachman, L. and Song, J. 2011. AdNext: a visit pattern-aware mobile advertising system for urban commercial complexes. *Proceedings of the 12th Workshop on Mobile Computing Systems and Applications*, ACM HotMobile 2011: 7–12.
- King, R.D., Oliver, J.R.S.G., Young, M., Aubrey, W., Byrne, E., Liakata, M., Markham, M., Pir, P., Soldatova, L.N., Sparkes, A., Whelan, K. E. and Clare, A. 2009. The Automation of Science. *Science* 324(5923): 85–89.
- King, R.D., Whelan, K.E., Jones, F.M., Reiser, P.G.K., Bryant, C.H., Muggleton, S.H., Kell, D.B. and Oliver, S.G. 2004. Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* 427: 247–252.
- Koller, D. and Friedman, N. 2009. *Probabilistic Graphical Models*. Cambridge, MA: MIT Press.
- Kosinski, M., Stillwell, D. and Graepel, T. 2013. Private traits and attributes are predictable from digital records of human behaviour. *Proceedings of the National Academy of Sciences (PNAS)* 110(15): 5802–5805.
- Kuhlmann, H. 2005. *Living Walden Two: B.F. Skinner's Behaviorist Utopia and Experimental Communities*. Urbana and Chicago, IL: University of Illinois Press.
- Lockerd-Thomaz, A. and Breazeal, C. 2008. Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence* 172(6-7): 716–737.
- Messing, S. and Westwood, S. J. 2012. How Social Media Introduces Biases in Selecting and Processing News Content. Working paper, available at <http://www.stanford.edu/seanjw/papers/SMH.pdf>
- Muir, L. 2012. Control Space? Cinematic Representations of Surveillance Space between Discipline and Control. *Surveillance & Society* 9(3): 263–279.
- Nilsson, N. J. 2010. *The Quest for Artificial Intelligence*. Cambridge: Cambridge University Press.
- Olsthoorn, P. 2012. De Almachtige Zoekmachine: Google van Online TomTom naar Personal Coach. In: *Voorgeprogrammeerd: Hoe Internet ons Leven Leidt*, eds van 't Hof, C., Timmer, J. and van Est, R. Rathenau: Boom Lemma uitgevers. In Dutch.
- Palmås, K. 2011. Predicting what You'll do Tomorrow: Panspective Surveillance and the Contemporary Corporation. *Surveillance & Society* 8(3): 338–354.
- Pariser, E. 2011. *The Filter Bubble*. London: Viking (Penguin Books).
- Poole, D. 2010. Probabilistic Programming Languages: Independent Choices and Deterministic Systems. In: *Heuristics, Probability and Causality: A Tribute to Judea Pearl*, eds Dechter, R., Geffner, H. and Halpern, J., 253–269. London: College Publications.
- Reichmann, W. J. 1961. *Use and Abuse of Statistics*. London: Methuen.
- Resnick, M. 1994. *Turtles, Termites, and Traffic Jams*. Cambridge, MA: MIT Press.
- Richards, N.M. 2013. The Perils of Social Reading. *Georgetown Law Journal* 101(3): 689–99.
- Rustin-Paschal, N. 2011. Online Behavioral Advertising and Deceptive Campaign Tactics: Policy Issues. *Wm. & Mary Bill Rts. Journal* 19: 908–925.
- Schermer, B.W. 2011. The Limits of Privacy in Automated Profiling and Data Mining. *Computer Law & Security Review* 27: 45–52.
- Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Ramones, S.M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., 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* 8(9): 1–15.
- Sharkey, N. 2008. 2084: Big robot is watching you – Report on the future of robots for policing, surveillance and security. Available at http://staffwww.dcs.shef.ac.uk/people/N.Sharkey/Future_robot_policing_report_Final.doc.
- Skinner, B.F. 2005[1948]. *Walden Two*. Indianapolis: Hackett Publishing Company Inc.
- Skinner, B.F. 1976. *About Behaviorism*. New York: Vintage Books.
- Steichen, B., Ashman, H. and Wade, V. 2012. A Comparative Survey of Personalised Information Retrieval and Adaptive Hypermedia Techniques. *Information Processing and Management* 48(4): 698–724.
- Tene, O. 2011. Privacy: The New Generations. *Int. Data Privacy Law* 1(1): 15–27.

- Tene, O. and Polonetsky, J. 2012. Big Data for All: Privacy and User Control in the Age of Analytics. *Northwestern Journal of Technology and Intellectual Property* 11(5): 239-273.
- Thurman, N. and Schifferes, S. 2012. The Future of Personalization at News Websites. *Journalism Studies* 13 (5): 775–790.
- Tocha, E., Wang, Y. and Cranor, L. F. 2012. Personalization and Privacy: A Survey of Privacy Risks and Remedies in Personalization-based Systems. *User Model User-Adap Inter* 22: 203–220.
- Turow, J. 2011. *The Daily You*. New Haven and London: Yale University Press.
- Tversky, A. and Kahneman, D. 1981. The Framing of Decisions and the Psychology of Choice. *Science* 211(4481): 453–458.
- Vaidhyanathan, S. 2011. *The Googlization of Everything*. Berkeley: University of California Press.
- Van den Broeck, G., Thon, I., van Otterlo, M. and De Raedt, L. 2010. DTProbLog: A Decision-Theoretic Probabilistic Prolog. *Proceedings of the National Conference on Artificial Intelligence (AAAI)* 2: 1217-1223.
- van Otterlo, M. 2009. *The Logic of Adaptive Behavior*. Amsterdam: IOS Press.
- van Otterlo, M. 2013. A Machine Learning View on Profiling. In: *Privacy, Due process and the Computational Turn – Philosophers of Law meet Philosophers of Technology*, eds M. Hildebrandt and K. de Vries, 41-64. Oxon and New York: Routledge.
- van 't Hof, C., Timmer, J. and van Est, R. eds. 2012. *Voorgeprogrammeerd: Hoe Internet ons Leven Leidt*. Rathenau: Boom Lemma uitgevers, in Dutch. (Translation: Preprogrammed: How internet rules our lives).
- Vanderbilt, T. 2013. The Future of Search. *Wired Magazine* (UK), January issue, 104–111.
- Whyte, J. 2004. *Crimes Against Logic*. New York: McGraw-Hill.
- Wiering, M.A. and van Otterlo, M. eds. 2012. *Reinforcement Learning: State-of-the-Art*. Heidelberg, New York, Dordrecht and London: Springer.
- Yang, Q. 2009. Activity Recognition: Linking Low-level Sensors to High-level Intelligence. *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence*, 20-25.
- Zarsky, T.Z. 2012. Automated Prediction, Perception, Law and Policy. *Communications of the ACM* 15(9): 33–35.
- Zwick, D. and Knott, J. D. 2009. Manufacturing Customers: The Database as New Means of Production. *Journal of Consumer Culture* 9: 221–247.