Chapter 10
The Ethics of Big Data Applications in the Consumer Sector

Markus Christen, Helene Blumer, Christian Hauser, and Markus Huppenbauer

Abstract Business applications relying on processing of large amounts of heterogeneous data (Big Data) are considered to be key drivers of innovation in the digital economy. However, these applications also pose ethical issues that may undermine the credibility of data-driven businesses. In our contribution, we discuss ethical problems that are associated with Big Data such as: How are core values like autonomy, privacy, and solidarity affected in a Big Data world? Are some data a public good? Or: Are we obliged to divulge personal data to a certain degree in order to make the society more secure or more efficient? We answer those questions by first outlining the ethical topics that are discussed in the scientific literature and the lay media using a bibliometric approach. Second, referring to the results of expert interviews and workshops with practitioners, we identify core norms and values affected by Big Data applications—autonomy, equality, fairness, freedom, privacy, property-rights, solidarity, and transparency—and outline how they are exemplified in examples of Big Data consumer applications, for example, in terms of informational self-determination, non-discrimination, or free opinion formation. Based on use cases such as personalized advertising, individual pricing, or credit risk management we discuss the process of balancing such values in order to identify legitimate, questionable, and unacceptable Big Data applications from an ethics point of view. We close with recommendations on how practitioners working in applied data science can deal with ethical issues of Big Data.

M. Christen (✉)
Center for Ethics, University of Zurich, Zurich, Switzerland
e-mail: christen@ethik.uzh.ch

H. Blumer · C. Hauser
Department of Entrepreneurial Management, University of Applied Sciences HTW Chur, Chur, Switzerland

M. Huppenbauer
Center for Religion, Economy and Politics, University of Zurich, Zurich, Switzerland

M. Braschler et al. (eds.), *Applied Data Science*,
https://doi.org/10.1007/978-3-030-11821-1_10

1 Introduction

Terms like "Big Data," "digitalization," or "Industry 4.0" have become keywords for indicating the radical changes implied by the pervasive use of digital technology. Big Data basically stands for the fact that today we are not only able to create, record, store, and analyze large amounts of heterogeneous data but that data about almost any fact in the world is available for such purposes as a commodity (Mayer-Schönberger and Cukier 2013). Computers, smartphones, and wearables, as well as the emerging "Internet of things" routinely generate digital data about where we are, what we do, and with whom we communicate. Actually, it is an inherent property of digital technology to generate data in order to function properly. For example, a telecommunication provider needs to "know" the geographic location of a smartphone for providing even its most basic functions (communication). The use of this technology generates data on processes that were mostly obscure in the "pre-digital age"—for example, if one compares rummaging in an old-fashioned book-store with searching for books on Amazon, where each click leaves a digital trace. But digitalization not only makes it easier to create data, it also has become increasingly cheaper and convenient to store and analyze it. Production and consumption processes thus become ascertainable in a way that was almost unthinkable a few decades ago.

Such radical changes spark both hopes and fears. Some believe that Big Data will be the "oil of the twenty-first century,"[1] that is, an enormous resource for innovation, progress, and wealth. Others consider Big Data to be a fundamental threat for freedom and privacy—a demonic instrument of an Orwellian surveillance state (Helbing et al. 2015). Both scenarios are probably overstated, but they point to difficult ethical problems that are associated with Big Data: How are core values like autonomy, privacy, and solidarity affected in a Big Data world? Are some data a public good? Are we obliged to divulge personal data to a certain degree in order to make society more secure or more efficient?

In this chapter, we will discuss these questions from the perspective of Data Science applications in the consumer sector. This concerns, for example, personalized advertising, tailored product offers, or individualized pricing. What are the ethical questions raised by such applications? Which values have to be weighed against each other in such cases? What are realistic chances and risks of Big Data for consumers? The answers to these questions given in this chapter rely on a study executed by the authors for the Swiss Academy of Engineering Sciences (Hauser et al. 2017). In the following Sect. 2, we provide some background information on ethical thinking in the field of Big Data and on methodological aspects. In Sect. 3, we outline the results of a bibliometric study and we describe five use cases of Big Data applications. In Sect. 4, we will evaluate those case studies from an ethical

[1]The notion of "data as the oil of the twenty-first century" first appeared in 2006 and has become a widespread quote for outlining the economic potential of Big Data; see https://www.quora.com/Who-should-get-credit-for-the-quote-data-is-the-new-oil (last accessed August 10, 2016).

perspective. This analysis will result in some “lessons learned” in Sect. 5, that is, in suggestions how practitioners working in applied data science can deal with ethical issues raised by Big Data. We close by conclusions regarding the possible role of the state in empowering companies and customers for dealing with the ethics of Big Data.

2 Background Information

2.1 *Big Data Ethics*

The ethical debate concerning Big Data is embedded in a broader discourse on the ethics of data protection that has been developed over several decades (Davis and Patterson 2012). As our bibliometric study indicates (Sect. 3.1), terminologies related to privacy and surveillance still dominate the ethics discourse with respect to Big Data. Certainly, depending on the domain of application, other values will be of importance as well. For example, in the insurance sector, we can expect that the value of solidarity will be of particular importance, since some Big Data applications in this industry may involve a significant potential for unjustified discrimination, and this could lead to a destruction of common grounds for society. Another relevant value affected by Big Data is fairness, as the current ecosystem around Big Data may create a new kind of digital divide: The Big Data rich and the Big Data poor—and, for example, large insurance companies may be representatives of former, putting them into a position to have privileged access to knowledge on societal processes (Boyd and Crawford 2012).

The large majority of papers published so far on the ethical debate on Big Data concern either the health sector or research ethics in the context of Big Data (to illustrate: 13 out of the 20 most cited papers in Scopus searched with the keywords “big data” and “ethic*” refer to healthcare issues; search dated 2016). This indicates a need for further developing the ethical discourse with respect to Big Data in other fields—including consumer applications.

Within the ethical discourse, some authors suggest to consider the fundamental societal changes implied by Big Data: A profound effect of the digitalization of information is that the boundaries around which human beings conceptualize and organize their social, institutional, legal, and moral world are compromised or relativized (van den Hoven et al. 2012). The traditional offline distinctions and demarcations of separate social spheres (family, work, politics, education, healthcare, etc.) are threatened by the enhanced reproducibility and transmissibility of digitalized data and the use of multidimensional Big Data analytics. Thus, a first line or research with respect to Big Data ethics concerns contextual integrity of social spheres as proposed by Helen Nissenbaum (2004), whereas spheres are defined through the expectations and behaviors of actors.

In this framework, what is often seen as a violation of privacy is oftentimes more adequately construed as the morally inappropriate transfer of personal data across the boundaries of what we intuitively think of as separate “spheres of justice” or “spheres of access” (van den Hoven 1999). This complex moral reason also accounts for the moral wrongness of discrimination. Discrimination of a person P in a particular context implies the use of information about P to her disadvantage while the information is morally irrelevant to that context (e.g., using the information about the gender of a person for determining his or her salary). The art of separation of spheres (or contexts of use) and the blocked exchanges would prevent discrimination. It secures that information is used in contexts where it is relevant and morally appropriate (van den Hoven et al. 2012).

We argue that spheres also differ with respect to the emphasis of certain values. For example, equality (the right of different groups of people to receive the same treatment with respect to the same interests irrespective of their social position) plays a particularly important role in the health sphere, fairness (treatment in accordance with accepted rules or procedures of justice) is an overarching value in the business domain and freedom (the power or right to act, speak, or think as one wants) is a guiding value in the political sphere. Related to this, Alan Fiske’s “social relational theory” proposes that there are various but universal types of social interactions or relationships, each of them describing qualitatively distinct structures with their own norms and rules of interactions (Fiske and Tetlock 1997). Within each type, people can usually make trade-offs without great difficulty, but between the domains, comparisons and, for example, applying market-price rules are problematic. We propose that a deeper understanding of the ethical issues raised by Big Data requires an analysis of which values are affected by Big Data applications and how the understanding and weight of these values depends on different social spheres or types of social relationships (Lane et al. 2014). Some of these values may have the status of “protected values” (Tanner and Medin 2004) for the involved persons, which further complicates the picture. Previous studies have shown that when people expect protected values to be under threat, they are likely to trigger reactions of outrage and objection to alleged violations (Tetlock et al. 2000).

A second line of research relevant for Big Data ethics concerns the question how ethics can be integrated into the design process of information technology. Creating Big Data applications is an issue of data product design—and making such a process compliant with ethical requirements puts engineers and managers in the focus. A frame of reference is the methodology of value sensitive design (VSD) that has been put forward by Batya Friedman et al. (2006). In her words, employing value-sensitive design means to account “[. . .] for human values in a principled and comprehensive manner throughout the design process.” Through a combination of conceptual, empirical, and technical investigations, one investigates how people are affected through the technology to be designed. Case studies demonstrating how this idea can be implemented in practice can be found in Friedman et al. (2006).

Today, a number of researchers have used the methodology of VSD and it also found its way into textbooks used in engineering education (Van de Poel and Royakkers 2011). Through a combination of conceptual, empirical, and technical investigations, VSD investigates how direct and indirect stakeholders are affected through the technology to be designed. VSD means choosing, among available technological solutions, those meeting normative requirements and desiderata. "Normative requirements" is a broader concept than requirements sanctioned by law. When a software architecture is designed, there is normally more than one way for the software to solve the problems that it is intended to solve; in making specific engineering choices at different levels of software design, developers implicitly or explicitly express their commitment to grounding principle and values, thereby attributing (a different) importance to them. In this approach, it is assumed that maximizing user satisfaction is not the only goal of good software design, because user satisfaction should not be achieved by sacrificing more important normative constraints.

2.2 Methodology of the Study

The study on which this chapter relies was based on a qualitative and quantitative literature analysis, on expert interviews, and on two workshops with practitioners (company representatives as well as data protection officers).

The literature analysis was performed in two scientific databases (Web of Science and Scopus)[2] as well as in the media database Factiva[3]; the timeframe was restricted to 2006–2015. The papers identified in this way served for a differentiation of the various thematic strains discussed in Big Data. Based on these papers, we identified keywords for characterizing Big Data publications that discuss ethical aspects along six categories: privacy, security, surveillance, harm, self-related topics, and ethics in general.[4] We also analyzed the disciplinary diversity of highly cited Big Data papers (those who received at least more than 10 citations until March 2016) by referring to the subject categories of the WoS papers (those categories refer to the discipline(s) to which the journal, in which a paper has been published, is attributed).

[2]Web of Science (WoS): https://apps.webofknowledge.com; Scopus: http://www.scopus.com. The search term was in both databases "big data" (WoS: search under "topics"; Scopus: search in the category "title, abstract, keywords").

[3]This database is hosted by Bloomberg and includes contributions from the most important international print media (such as New York Times, etc.), and contributions from a multitude of information sources mostly from the business domain; see: https://global.factiva.com. The search term was "big data" as well.

[4]Each ethics category was characterized by a set of 2–5 keywords as follows; the specificity of each keyword was checked individually: Privacy (privacy OR anonym*), security (security OR protection), surveillance (surveillance OR profiling), harm (discrimination OR harm), self-related (identity OR reputation OR ownership) ethics in general (ethic* OR moral OR fairness OR justice OR autonomy). For the quantitative analysis, these keyword sets were combined with "big data" using the Boolean operator AND. Those keywords had to be present either in the title, the abstract, or the keywords of the scientific papers. Those categories do not match the eight values identified further below in the paper, because some of them such as contextual integrity are hard to quantify using a bibliometric approach.

In the two workshops, 22 experts from Swiss institutions and companies were present in total. In the workshops, current and likely future Big Data applications were identified and discussed with respect to the associated ethical questions. The experts emerged from the following industries: banking, consultancy, insurances, marketing, retail business, soft- and hardware producers, telecommunication, and transport. In addition, cantonal (state-level) and federal data protection officers and scientists active in the field complemented the workshop participants. The experts identified five paradigmatic use cases that are discussed in Sect. 3:

– Prevent debt loss
– Improve risk management
– Tailor offer conditions
– Increase the efficiency of advertising
– Create business innovations

They also pointed to eight groups of ethical values that are affected by Big Data applications:

– Privacy protection
– Equality and non-discrimination
– Informational self-determination
– Controlling the own identity
– Transparency
– Solidarity
– Contextual integrity
– Property and copyright

Those eight value groups will serve as a framework for the ethical assessment in Sect. 4.

3 Big Data in the Scientific Literature and in Business

3.1 Bibliometric Study

The bibliometric analysis serves to provide a first overview on the topic of Big Data by showing the frequency of published papers that contain the keyword "big data", the relative weight of the six ethics categories used in the bibliometric study, and the disciplinary spectrum of highly cited papers. The original study has been performed in March 2016 but has been updated in February 2018, leading to some changes as discussed in the text. Figure 10.1 shows the frequency of Big Data articles both in the scientific literature as well as in the media, that is, the number of papers published in a single year compared to all papers found in the reference timeframe. It is striking that almost no paper has been published before 2011—the first paper that uses the term "big data" in the current understanding was published in 1998. Since 2011, an enormous growth in the literature can be observed, whereas the growth rate was

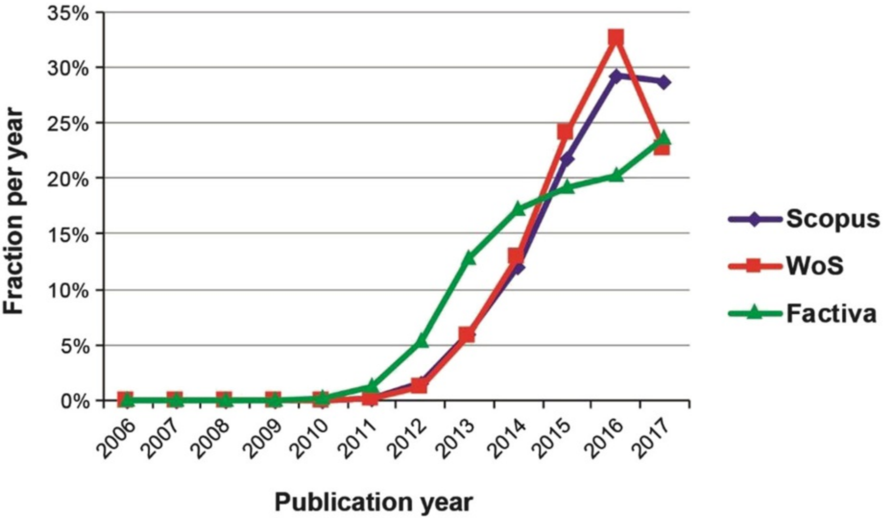

Fig. 10.1 Per-year-fraction of Big Data papers published between 2006 and 2017 in each database; the data of 2017 are incomplete, that is, the diminishment in this year is likely an artefact

higher in the general media compared to the scientific literature: in Factiva, the trend seems to weaken to some degree, but in the scientific literature, more than half (58.1% WoS and 55.4% Scopus) of the papers have been published in the last 2 years. The diminishment in the last year is mainly an effect of database curation, as both databases do not yet contain all papers published in 2017 at the time of updating (February 21, 2018).[5]

The enormous dynamics in the field is also discernible when looking at the number of highly cited papers (>10 citations per paper): In 2016, we identified 164 highly cited papers (2.60% of all papers identified), in 2018, 1333 papers (5.79%) fulfilled this criterion.

We remind that this search only focuses on the use of the term “big data” in the literature, not on more generic topics within computer science that are today attributed to the field of Big Data (such as data mining or data visualization techniques). Thus, the data does not allow to make inferences on how those scientific fields have developed over time.

Figure 10.2 shows the relative weight of Big Data papers that contain keywords of one of the six ethical categories (see Footnote 4). An interesting observation is that in the lay media (Factiva), papers referring to security and “self” (reputation and ownership issues) have a much higher weight compared to the two scientific

[5]Database curation also affects the comparison of the results from 2016 with those from 2018: We found that the number of entries already in the year 2015 more than doubled in WoS, but was comparable for the earlier years.

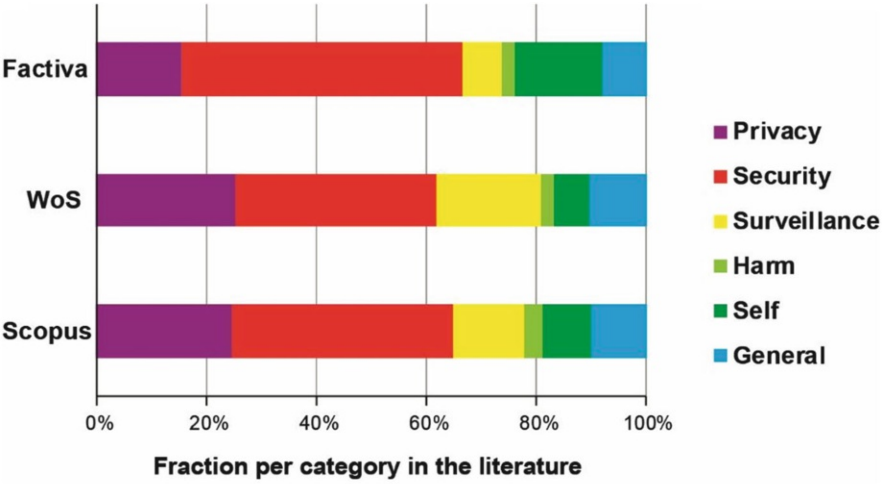

Fig. 10.2 Fraction of papers referring to one of the six ethics categories published between 2006 and 2017 in each database

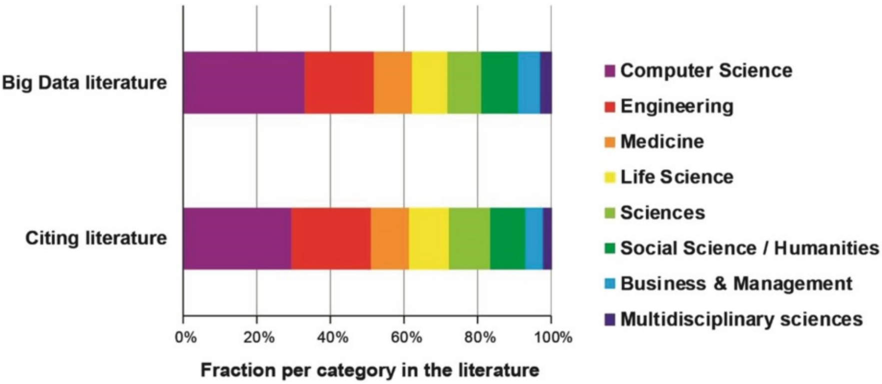

Fig. 10.3 Disciplinary profile of highly cited Big Data papers compared to the papers that cite them; 2018 data

databases, where the "classic" topics of computer ethics, namely, privacy and surveillance, are more relevant. This pattern did not change in the 2018 update.

Finally, Fig. 10.3 outlines the disciplinary profile of the highly cited papers (1333 papers) and compares it with the profile of the papers citing them. This analysis gives an estimation on the "knowledge flow," that is, which disciplines tend to cite papers more (or less) often. Here, interesting differences between the 2016 and 2018 data is discernible. In 2016, in particular humanities and social sciences (the fraction increased from 6.6% to 13.2%), life sciences (increase from 12.2% to 16.5%) cite Big Data papers more often, indicating that the debate is more pronounced in these disciplines. In the 2018 data, such a difference when comparing publications and citations is not discernible any more. Humanities and social sciences, for example, now account for 10.0% of all publications compared to 6.6% in 2016. Both in the 2016 and 2018 data, a decrease in the domain "economy/management" is visible (from 6.9% to 4.4% in 2016 and from 6.2% to 4.8% in 2018), which could indicate a less-developed discussion of Big Data in this domain compared to other disciplines.

In summary, the bibliometric analysis shows that the topic "Big Data" is very young and displays indications of a "hype." Furthermore, there are different weights with respect to the ethical debate when comparing the scientific literature with the lay media. Finally, there are indications that Big Data is particularly discussed in the life sciences, humanities, and social sciences.

3.2 Use Cases

In the following, we briefly describe five use cases for outlining Big Data applications in the consumer sector. In Sect. 4, we discuss ethical issues of Big Data based on these cases.

Case 1: Prevent Debt Loss Big Data allows new ways for companies to assess the payment moral of their customers based on the customers' digital behavior. Traditionally, companies relied on registries such as the Swiss "Zentralstelle für Kreditinformation" or the German "Schutzgemeinschaft für allgemeine Kreditsicherung" for credit rating. As an alternative, social scoring based on Big Data is increasingly used. For this, algorithms analyze entries and behavior of customers on social networks (friends, likes, leisure activities) as well as information that—on the first sight—seems to be unrelated to credit rating such as search behavior, fonts used when writing, speed of filling out forms, or technical data of the computer used when surfing the Internet. Social scoring is quite common in online shopping, for example, for determining whether a customer is allowed to pay the bill using an invoice. Also in banking, social scoring gains importance. For example, some banks provide credit under the condition that the customer downloads an App that collects personal information of various kinds (geographic location, duration of phone calls, address book information, etc.)—the more data the customer reveals, the higher are the chances for getting a credit (an example of such a social scoring system for improving the access to credits is provided by the Australian loans and deposits application platform Lodex; Graham 2017).

Case 2: Improve Risk Management For many industries such as insurances, risk management is key for business success. Big Data provides the opportunity to better evaluate risks, for example, the probability and magnitude of damages. In particular, risks can be assessed more individually. An example is the use of telematics solutions in car insurances. Information on how the customer is driving (speed, acceleration, braking behavior, duration of driving, distances, etc.) allows to calculate the probability of an accident or of car theft, leading to "pay as you drive" models. Wearables such as smartwatches or fitness trackers provide information that is relevant for health insurances. For example, unhealthy behavior of customers can be tracked more easily, allowing for prevention to decrease the insurance rate. Also, the genetic information of people can be more easily determined and shared using online services. Big Data also allows for better identification of fraud by customers using profiling and predictive modelling.

Case 3: Tailor Offer Conditions Traditionally, companies calculate prices for large customer groups based on costs, prices of competitors, and aggregated customer behavior. Big Data now allows in principle to determine the individual "best price" for each customer. The first step for this is dynamic pricing, which has become a standard in several industries. Airlines, e-businesses, or gas station providers use algorithms to dynamically change prices based on various types of information (demand, availability, weather, time in the day, behavior of competitors). Individualized prices are the next step, allowing to best skim the consumer surplus. For this, not only environmental factors used in dynamic pricing but also information on the individual customer is used (e.g., gender, age, geographic origin, friends, personal preferences) based on cookies, customer cards, smartphone ID, GPS position, IP address or other technical means. For example, an online tour operator displayed higher prices to Apple users, because they tend to be less price-sensitive when booking (Mattioli 2012). An online shop gave away discount tickets based on the individual shopping behavior of the customer for nudging the customer to products with higher prices (Metzler 2016).

Case 4: Increase the Efficiency of Advertising Advertising is effective when it reaches the customer who is potentially interested in the product—but traditional advertising campaigns mostly work based on the "shotgun approach" (e.g., billboard advertising). The accuracy of such campaigns is increased using Big Data, based on search history, social network data, GPS data, etc., of customers. A common technique is re-targeting: a customer that searched for a specific product finds advertising of this product later on many other sites he/she is visiting. Pre-targeting aims to show potential products to the customer based on his/her online behavior. Geo-targeting aims to show advertising related to the geographic localization of the customer (e.g., a nearby restaurant). Future applications that are currently investigated experimentally in computer games include emotional targeting: based on visual (face expression) and auditory information (voice recognition), the emotional state of the customer is assessed in order to display advertising adapted to his/her emotional state.

Case 5: Create Business Innovations Big Data is also used for generating new revenue sources or enlarging the product or service portfolio of a company. For example, companies could sell the information of their customers to other companies, which includes the possibility to supplement existing products (e.g., sportswear) with sensors (wearables) that generate this data. Telematics systems in cars can be used to better identify bugs in new models or provide new maintenance services. Data emerging from social networks, etc., can be used to identify trends and adapt product developments to such trends (e.g., in car manufacturing or streaming services that develop own TV shows). Data generated from voice recognition software can be used to increase voice control technology or related services such as translation. Big Data also allows for innovations in infrastructure or traffic planning. For example, traffic jams could be prevented by such systems.

4 Ethical Evaluation

In the context of economy, ethics is usually construed to be an antagonist of business in the sense that ethics defines which business activities are legitimate and which are not. Although this is one of the functions of ethics, such a perspective is missing the following aspects:

1. Market economy itself has a moral foundation by assuming that a regulated market economy with informed actors serves to pursue ethical goals such as individual freedom and public welfare.
2. Ethical values and norms are usually abstract and need to be applied to concrete problems, which is in most cases not a straightforward process.
3. The claims associated with ethical values and norms can be in conflict to each other, which requires some degree of balancing.

The following analysis structured along eight ethical values is based on these preconditions.

4.1 Protection of Privacy

Article 12 of the Universal Declaration of Human Rights[6] declares that “no one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation. Everyone has the right to the protection of the law against such interference or attacks.” The goal of this norm is to protect spheres of life of the individual where he/she can move, develop, and behave freely. In the current Swiss data protection law[7] (which is currently under revision), this value of privacy is protected by the two principles of “purpose limitation” (“Zweckbindung”) and “data minimization” (“Datensparsamkeit”): data should be collected only for specific purposes and only as much as is needed for this purpose. All five use cases above involve the potential to infringe these principles. For example, smartphone apps may violate these principles by collecting data that is

[6]Available at: http://www.un.org/en/universal-declaration-human-rights/ (last access: February 28, 2018).

[7]Available at: https://www.admin.ch/opc/de/classified-compilation/19920153/index.html (last access: February 28, 2018).

not necessary for their primary purpose. These risks increase when data emerging from different sources are combined in order to gain new insights, for example, in the case of pre-targeting (Case 4). However, the potential of Big Data lies in this combination of various data sets. Obviously, secondary use and recombination is in conflict with the principle of purpose limitation—and anonymization is not sufficient given the potential of re-identification when integrating data from different sources.

However, in the case of preventing debt loss (Case 1) and increase of risk management (Case 2), the violation of privacy has to be balanced with legitimate claims of the companies, which have the right to know about the solvency of their debtors and other business-relevant risks. Thus, not every violation of privacy is equally ethically problematic—and in cases 1 and 2 one would have to look at the individual circumstances. However, when in the case of tailoring offer conditions (Case 3) data of very different type are integrated (financial transactions, credit rating, medical treatments, social relationships, etc.), an unjustified violation of privacy is likely. Furthermore, missing transparency, the potential of discrimination and the violation of contextual integrity of the data are additional risks that often go hand in hand with such applications; they are discussed below.

Increasing the efficiency of advertising (Case 4) and creating business innovations (Case 5) are legitimate goals of companies. In those cases, one has to balance the gains for the customers (e.g., not being disturbed by advertising that is not interesting at all for the customer) with the drawbacks (e.g., surveillance by the company). From an ethical perspective, violation of the privacy of the customer seems to be justified, if he/she is informed on the magnitude of data collection, if he/she has consented to data collection, and if there is a realistic alternative when not consenting. The problem that the general terms and conditions of companies are often hard to understand is discussed further below in Sect. 4.5.

4.2 *Equality and Non-discrimination*

Discrimination means the unequal treatment of persons that is not justified by ethically sound reasons. Non-discrimination is founded by fairness intuitions that are undisputed and that are mirrored by legal principles such as equality before the law. However, non-equal treatment of people is not necessarily ethically problematic or may even be requested (e.g., differences in wages when the work performance of people differs). But when non-equal treatment is based on criteria (such as gender, race, or political opinions) that are not relevant for accessing certain goods or positions, it becomes ethically problematic discrimination.

This can happen when tailoring offer conditions (Case 3): algorithms may classify persons based on properties they themselves can hardly influence. This may be the reason that businesses currently tend not to use individualized pricing directly but rely on mechanisms such as discount tickets—although this is not fundamentally different from the former. Using Big Data for tailoring offer conditions involves a systematic non-equal treatment of customers. However, it is in principle not ethically wrong that the willingness to pay of a customer is part of the pricing mechanism, as long as one does not exploit a state of emergency and as long as there is no monopoly. Economic institutions such as the bazar or auctions are accepted mechanisms of using the willingness to pay for the pricing mechanism. The problem of Big Data, however, is that information asymmetry is involved between the vendor and the customer, for example, when the company has access to information about the psychological profile of the customer based on his or her behavior on social networks and the customer is unaware that this data (and the associated model) is available to the vendor. This may undermine the mechanisms of efficient allocation in markets. Of particular ethical relevance is that customers may not know based on which mechanisms they may be treated.

Individualized prices are problematic when they are used ubiquitously within an industry and customers have no possibility for evasion. However, this argument is only valid when there is only one mechanism to determine the individual price. This would require that all providers use the same algorithm and the same data—the situation of a classic monopoly or syndicate, that is, a market without competition.

The example of social scoring (Case 1) involves both benefits and risks: On the one hand, this mechanism offers access to credits for people who usually would never be able to enter the classical banking system. On the other hand, those digital credit providers require disclosing sensitive personal data that is not the case in classical banking. Is this unjustified discrimination? Again, this example requires balancing of the violation of privacy with the right of the company to prevent credit losses. Ethically problematic is when the disclosed data allows for inferences for third persons or is later used in a way that violates the contextual integrity of the data.

4.3 Informational Self-Determination

Usually, informational self-determination is defined as the right of an individual to decide upon collecting, storing, using, and transferring of personal data. This right is founded in the value of autonomy, that is, the ability to shape and control the personal life. A practical expression of informational self-determination is informed consent—a concept that originally emerges from the medical sphere. In the case of data this means that a person should consent explicitly and in an informed way to the use of his/her personal data.

Informational self-determination is likely to be violated when targeted advertising aims to manipulate the person, for example, in the case of emotional targeting (Case 4). Problematic is that the persons did not explicitly consent to the use of their data (e.g., facial expressions). Furthermore, the manipulative nature of the intervention may undermine the process of free opinion forming. That advertising has some manipulative character is not new and willful deception is surely wrong. But the use of Big Data has the potential to strongly increase the efficiency of such mechanisms. A general statement is, however, not possible and requires a case-by-case evaluation. Another problem is that customers who insist on their right of informational self-determination may not be able to get certain services when denying access to personal information or would have to pay substantially more. This may result in a de facto discrimination that can be considered unfair.

4.4 Controlling the (Digital) Identity

Closely related to informational self-determination is the claim of being able to control the own digital identity. Digital identities can be the result of Big Data applications, when features of customers are aggregated and correlated. At first glance, single data points such as how people use a keyboard or when they usually do phone calls seem to be unrelated to, for example, the credit rating of a person. However, when multidimensional data is aggregated and analyzed, categories can be created in order to match the digital identity of persons with these categories—as exemplified in Case 1. This is particularly ethically problematic, when the person does not know that his/her data is used in this way and when the person has no possibility to change his/her attribution to one of these categories, in particular when the attribution is obviously mistaken or was based on a spurious correlation.

An additional problem is that such digital profiles may include outdated data—so there is no forgetting or prescription. Data related to personal situations are context-dependent with respect to the age of the person: youthful folly leave digital traces that then may fall back to the adult person. Companies that use automatized techniques of Big Data analysis without being aware of this time- and context-dependency of personal data treat their customers unfairly, because they cannot contribute to the interpretation of their own data.

However, the fact that digital identities are always incomplete and selective is not per se ethically problematic, as it is in the nature of things. Persons themselves often construe digital identities adapted to contexts (e.g., a profile on a dating network compared to a profile on a business network), which is actually an expression of informational self-determination, as long as there is no intention of deception. Rather, the incomplete digital identities that companies may have from their customers result from the principle of data minimization.

Furthermore, if people change their digital behavior in anticipating that the data generated through this behavior helps to construct digital identities is not a new problem. Customers, for example, may change their online behavior in order to profit from discounts. This kind of heteronomy of the own digital identity is part of the way people tend to interact socially in order to increase personal advantages. An ethical problem, however, arises when no room for "non-strategic behavior" is left due to ubiquitous commercial surveillance of the digital behavior of people.

4.5 Transparency

Transparency in the context of business means that customers, business-partners, or investors are informed properly of the state of a company, its business processes, services, and products such that they can form an informed decision (e.g., whether to invest or not). Thus, transparency is also a precondition for any informed consent for using personal data. Without transparency, markets cannot function properly. In the case of Big Data, transparency means that every person has the right to know which data are collected about him/her and how they are used.

An obvious problem of transparency is that companies tend to provide extensive and incomprehensible terms and condition forms toward their customers. Although one can expect that customers should read these forms, the way they are presented practically prohibits an informed decision unless one is a legal expert. Furthermore, there is a lack of transparency which data are collected on which online platforms, who is analyzing this data, and to whom the results of these analyses are given. Often, app providers sell the collected data to third parties without informing their customer properly—they probably fear that customers will not use their services any more if they would know.

When tailoring offer conditions (Case 3), the algorithms used for generating the prices are usually confidential. Furthermore, the data cannot be checked with respect to their quality, reliability, and completeness. Whereas online trading made markets more transparent, because comparing prices became easier, Big Data now undermines this gained transparency. Thus, tensions appear between the claims of companies to protect their algorithms (i.e., the intellectual property associated with them) and the claim of customers for transparency. In liberal societies, solving this challenge is a task of the company primarily. Legal regulation should be considered when they fail doing so.

Using learning systems aggravates this problem: Those algorithms may increasingly be used for preventing debt losses. For example, so-called deep learning algorithms may learn classifications of risk ratings that are even intransparent for the software developers—they are "black boxes" (Pasquale 2015). This is a fundamental problem of many currently used machine-learning systems relying on neural networks, as there are currently few underlying theories that explain how or why the models are effective for a particular type of problem and no baseline exists to predict their eventual performance. Machine-learning models are equations that have no obvious underlying physical or logical basis. Reading these models provides no insight into the underlying phenomena, where they originated, or how they will behave in a given situation. Furthermore, a model may produce radically different results for two scenarios that seem quite similar to humans. This poses significant problems related to testing (and trusting) such algorithms (Informatics Europe & EUACM 2018).

This lack in transparency endangers the value of informational self-determination, because people may agree to reveal information that is processed in a way that leads to a new type of classification scheme where no person reasonably expects that this scheme emerges. The decision to reveal this information is thus based on the wrong assumption that one understands what one can reasonably infer from the information one discloses. For example, people may accept to disclose their favorite color, preferred food, and most-liked movies on a convenient website by considering this information as unproblematic—and neither the person nor the provider of this website initially had the idea that a complex evaluation algorithm could infer out of this information the risk for insurance fraud. Furthermore, it could involve legal risk with respect to the new EU data protection legislation that restricts what the EU calls "automated individual decision-making"—the task of supervised machine learning like deep learning neural networks (Goodman and Flaxman 2016).

4.6 Solidarity

Solidarity concerns duties of mutual support in a community. In today's social state models, solidarity involves financial support in case of illness or poverty, based on the intuition that every human being could end up in a situation of need independent of negligence. In this way, solidarity provides the moral foundation of any type of insurance, whereas the range of solidarity is limited to some degree by the costs-by-cause principle. Persons who intentionally cause harm to themselves are less likely to benefit from the solidarity of others.

The challenge of Big Data is that the aggregation of multidimensional data for increasing risk management (Case 2) could lead to an extension of the costs-by-cause principle. A certain online behavior could be coupled to a higher risk of, for example, liability—making this behavior object of potentially "causing" liability, as the behavior results from a free choice. For example, playing certain video games may be correlated with a higher incidence of being absent from work due to illness—and the "choice" of playing these games may finally become a reason to deny solidarity in that case. These kinds of analyses could also provide a conceptual basis for prevention programs, for example, insurance companies could demand for certain diets or fitness programs to decrease certain health risks. This type of behavior control that is economically attractive for insurance companies is in conflict with the right of self-determination. To prevent this undermining of solidarity, there are legal barriers. In Switzerland, for example, health insurance companies are not allowed to exclude anyone from basic health insurance based on their behavior.

4.7 Contextual Integrity

The human environment is structured in social spheres that provide important reference points for human beings. They expect to be treated differently in a family context compared to, for example, in a governmental organization. They accept inequality in treatment in the economic sphere that they would not accept in the health, legal, or education spheres. The interpretation of moral values such as justice[8] or autonomy, and the rules related to these values differ along these social spheres. Accordingly, also the way information is produced and the social meaning people attach to information differs in these spheres. This is what is called the contextual integrity of information (see also Sect. 2.1). For example, if a person discloses personal information in the health sphere for research purposes, the moral foundation of this choice is to help other people. But if this information is used in a different sphere such as the economic sphere, to tailor offer conditions (Case 2) or to maximize profit (Case 5), the original intention to disclose this information and thus its contextual integrity is violated.

Based on these considerations, Big Data relying on multidimensional sources inherently entails the danger to violate contextual integrity of data. As data are increasingly traded by data brokers and are used in complex algorithmic or statistical models to make inferences on person groups, a violation of contextual integrity of data is hard to detect even for the commercial user of such data. This also undermines the value of transparency.

It is hard to evaluate which violations of contextual integrity pose an ethical problem, as the borders between social spheres as well as the rules within those spheres are fluid to some degree. The interpretation of values can change, for example, when individuals generally tend to disclose more personal information in social networks and also have this expectation toward their fellow humans—and in this way change the normative weight of privacy. Nevertheless, social spheres remain central points of reference for understanding the world, which explains why most people are filled with indignation when information emerging from their personal friends is used for individualized prices (Case 3).

4.8 Property and Copyrights

The functioning of the economic sphere depends on certain moral foundations, among which are the property right and the copyright. Both values are protected, for example, by the Swiss constitution. In the case of Big Data the question emerges, whether data also falls under these legal norms.

Using online services often entails the generation or disclosure of personal data, which is the basis for new revenue sources of companies (Case 5). The economic value of some companies is even measured based on the number of their customers and the amount of data they generate. This poses the question: who owns this data? When customers generate data on location or device usage when using smartphones, tablets, or computers: are these data streams creations in the sense of copyright law? Or is this the case not until companies use technologies to analyze this data?

[8] In the case of justice, different allocation rules exist. Examples include “an equal share for everyone” or “sharing according to needs”.

Depending on how these open questions are answered, the foundations of business cases of many companies active in Big Data may be shattered. For example, customers could have a share in the profits made by these data or they should have the right that the company deletes all personal data of this person, as foreseen by the new EU General Data Protection Regulation (GDPR).

From an ethical perspective, both the data provider and the companies that invest in the analysis of this data should have a fair share of the profit. For the latter, the current economic system is concerned, as companies only invest when they reasonably expect revenues—and they are free in investing or not. However, this is not the case for the data providers. The current model is that the customers are compensated by freely using certain services. In the future, this may not be sufficient and companies should start considering alternative compensation models. As the GDPR[9] foresees mechanisms such as data access (Art 15) and data portability (Art. 20; data subject have the right to receive the personal data concerning him or her), companies have incentives to increase trust and fairness toward their customers providing data. This may provide an incentive for new compensation models in order to avoid costly executions of those rights.

5 Lessons Learned

As our use cases demonstrate, Big Data is transforming the way companies develop, produce, offer, and advertise products and services. This may lead to added values both for businesses and their customers—but these applications entail also ethical risks, as they affect core values such as informational self-determination, transparency, and contextual integrity. Companies are well advised to be sensible to those risks, as customers and other stakeholders are likely to become more critical with respect to violation of those values. Also, Big Data applications need a "license to operate" and companies have to demonstrate a responsible use of data. We therefore suggest the following:

– **Take the "ethics case" into account:** When evaluating novel Big Data applications, companies should not only focus on the business case, but they should from the beginning also evaluate which of the core values described in this study may be affected in what way. This systematic involvement of ethics should be mapped on the appropriate corporate responsibility structures. Companies may also consider creating industry guidelines for a responsible use of Big Data.
– **Take the customer-point-of-view:** A simple test case for assessing the ethical risks of a Big Data application is the following: Would the customer still agree on disclosing his/her data when he/she knows exactly what is done with the data? What is the actual state of knowledge of the customers with respect to Big Data

[9]Available at: https://gdpr-info.eu/ (last accessed February 28, 2018).

and how likely is it that this will change? What are likely benefits customers would accept for trading in their data?

– **Create transparency and freedom to choose:** Trust and acceptance of consumers is a mandatory requirement for the successful application of Big Data. This requires that companies inform transparently and comprehensibly on how data is collected and used. Depending on the service, opt-in solutions and the provision of acceptable alternatives are successful strategies in that respect.

6 Conclusions

This chapter provided a summary of a study that intends to outline the ethics of Big Data applications in the consumer sector. It made clear that an ethical evaluation always involves a balancing in the single case in order to evaluate whether the violation of some core values can be justified. Individual companies may be overburdened in performing such an evaluation, making a public-private partnership advisable. The state should support the relevant industries in creating industry standards. In particular, some kind of standardization of general terms and conditions forms may be advisable in order to increase the informed consent capacity of customers. The goal should be that customers, and citizens, are empowered to better understand the way and magnitude of data collection and analysis in the age of Big Data.

References

Boyd, D., & Crawford, K. (2012). Critical questions for big data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication and Society, 15*(5), 662–679.

Davis, K., & Patterson, D. (2012). *Ethics of big data*. Sebastopol, CA: O'Reilly Media.

Fiske, A. P., & Tetlock, P. E. (1997). Taboo trade-offs: Reactions to transactions that transgress the spheres of justice. *Political Psychology, 18*, 255–297.

Friedman, B., Kahn, P. H., Jr., & Borning, A. (2006). Value sensitive design and information systems. In P. Zhang & D. Galletta (Eds.), *Human-computer interaction in management information systems: Foundations* (pp. 348–372). New York: M.E. Sharpe.

Goodman, B., & Flaxman, S. (2016). EU regulations on algorithmic decision-making and a "right to explanation". *2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*, New York, NY. Available at http://arxiv.org/pdf/1606.08813v1.pdf

Graham, B. (2017). *How banks could use an online 'score' to judge you*. news.com.au, November 6 2017. Available at http://www.news.com.au/finance/business/banking/how-banks-could-use-an-online-score-to-judge-you/news-story/009ca6df681c5fc69f583c4feac718c2

Hauser, C., Blumer, H., Christen, M., Huppenbauer, M., Hilty, L., & Kaiser, T. (2017). Ethical challenges of big data. *SATW Expertenbericht*. Available at https://www.satw.ch/digitalisierung/detail/publication/ethische-herausforderung-von-big-data/

Helbing, D., Frey, B.S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoven, J., Zicari, R.V., & Zwitter, A. (2015). Das Digitale Manifest. Digitale Demokratie statt

Datendiktatur. *Spektrum der Wissenschaft*, 17.12.2015. Last accessed August 10, 2016, from http://www.spektrum.de/news/wie-algorithmen-und-big-data-unsere-zukunft-bestimmen/1375933

Informatics Europe & EUACM. (2018): *When computers decide: European recommendations on machine-learned automated decision making*. Last accessed February 28, 2018, from http://www.informatics-europe.org/component/phocadownload/category/10-reports.html?download=74:automated-decision-making-report

Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (Eds.). (2014). *Privacy, big data, and the Public good: Frameworks for engagement*. Cambridge: Cambridge University Press.

Mattioli, D. (2012, August 23). On Orbitz, Mac users steered to pricier hotels. *The Wall Street Journal*. Available at http://www.wsj.com/articles/SB10001424052702304458604577488822667325882

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: Die revolution, die unser Leben verändern wird*. München: Redline.

Metzler, M. (2016, October 23). *Reiche bezahlen mehr.* NZZ am Sonntag. Available at http://www.nzz.ch/nzzas/nzz-am-sonntag/personalisierte-preise-reiche-bezahlen-mehr-ld.123606

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review, 79*, 119–157.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Cambridge: Harvard University Press.

Tanner, C., & Medin, D. L. (2004). Protected values: No omission bias and no framing effects. *Psychonomic Bulletin and Review, 11*(1), 185–191.

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M., & Lerner, J. S. (2000). The psychology of the unthinkable. Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, (5), 853–870.

Van de Poel, I., & Royakkers, L. (2011). *Ethics, technology, and engineering: An introduction*. Hoboken: Wiley.

Van den Hoven, J. (1999). Privacy and the varieties of informational wrongdoing. *Australian Journal of Professional and Applied Ethics, 1*, 30–44.

Van den Hoven, J., Helbing, D., Pedreschi, D., Domingo-Ferrer, J., Gianotti, F., & Christen, M. (2012). FuturICT – The road towards ethical ICT. *European Physical Journal – Special Topics, 214*, 153–181.