Workshop and lecture series in “Law & Economics of Innovation”
University of Zurich, Winter Term 2014

Prof. Dr. Andreas Heinemann
Prof. Dr. Stefan Bechtold (ETH)
Prof. Dr. Gérard Hertig (ETH)
J.D. and PhD Daniel Chen (ETH)

Comment

on

Tal Zarsky, The Privacy/Innovation Conundrum

by

[redacted]

Master of Law – University of Zürich

November 3, 2014

Introduction

After defining the meaning of the terms “privacy” and “innovation” in the context of his paper, Professor Tal Zarsky outlines five possible ways of linkage between the notions of privacy and innovation. A closer analysis of them, according to the author, eventually leads to the conclusion that strong privacy rules do not enhance the development of innovation in the same pace as they would in surroundings of free flow of data. In surroundings with no, or rather basic privacy regulations, firms would focus on enhancing their services or products rather than trying to match their innovative ideas with regulation standards. Thus, the compliance with privacy regulations would significantly encumber the drive to freely innovate, eventually depriving innovation from efficiency. As the paper does not outline possible ways how firms potentially violate a person’s privacy rights, the comment will in a first section mention a few methods of data collection and use in online advertisement and business operations. The second part is devoted to a brief examination on why individuals – despite the privacy threats lurking especially in the ICT realm – do not refrain from using products and services that are collecting their data, but rather use them to ever-higher extents. Part three of the comment will examine the counter-argument that privacy regulations do not hinder innovation and that therefore innovation is possible without jeopardizing privacy by using so called “Privacy by Design” instruments.

Privacy Related Use of Data

In the industrial revolution, labour, land and money were the motors for fundamental economic and social change.[1] Today, economy and society face another phase of fundamental changes due to a technological or informational revolution. Profit maximization is achieved though the access to a vast amount of information and the ability to process this complexity of information, thus monetizing it.[2] Societies connected to the Internet now produce every two days as much data as humanity did from the beginning of the distribution of data

[1] POLANYI, p. 48.

[2] CASTELLS, p. 14 ff.

though letterpress printing until the year 2003.[3] As Zarsky states it in the introduction chapter, we unarguably stand in the age of "Big Data". It has never before been easier (especially in the meaning of "cheaper" and "less labour-intense") for firms to obtain data from their customers that they, or other firms that purchase this (analyzed) data, could eventually use to enhance their performance and products by identifying patterns, making connections, predicting behavior and personalize interactions with their customers.[4] Earlier debates on privacy focused on the government collection and use of personal data, as well as, on the topic of surveillance with mainly the aim to prevent and detect crime. For non-governmental entities, the collection of personal data was too costly, thus it was mainly information on public figures that – by publishing this information – could infringe privacy rights.[5] Today in times of digitization however, with the facilitation of collection and storage of data, privacy related concerns do not only occur when it comes to the core values of rights regarding privacy (like name and image), but, as Zarsky states in Chapter III.2, also in situations, where the periphery of privacy is touched. According to Zarsky, this is due to the "extensive uncertainty" and "unnecessary broadness" of privacy laws. Regardless of the possible reasons, concerns regarding the periphery of privacy arise, when the data in question falls into the category of personal information (or so called "PII" – personally identifiable information). Despite possible jurisdictional differences, personal information in general can be defined as "any information, recorded or otherwise collected, relating to an identifiable individual" through direct, indirect or manipulated data linkage.[6] Of course, only in specific context, data can be considered personal and private. Thus, seemingly pure geographical, biological, transactional, historical, computational, etc. data may become personal when being linked to an identifiable individual.

[3] CAVOUKIAN et al, p. 5.

[4] GOLDFARB/TUCKER, p. 3f.; CAVOUKIAN et al, p. 7.

[5] GOLDFARB/TUCKER, p. 4; see also STRAHILEVITZ, p. 2013 for today's public figure privacy issues.

[6] CAVOUKIAN et al, p. 8.

Use of Date in Online Advertising and Business Operations

Online advertising firms mostly collect data for targetability and mesurability reasons.

In order to find out which kind of customer is most likely to be influenced best by what kind of ad, advertising firms expose a particular ad to a particular subset of potential ad target audience. For example, an online ad about e-bikes would only be shown to people that recently browsed websites for e-bike ratings and reviews.[7] Media platforms therefore need to get comprehensive data on the subset's recent browsing behavior.[8] The collection of such data is typically conducted by website relations through cookies, flash cookies or webbugs that enable media platforms to track relevant users over time and across websites. Surf- and clicktracking can then help advertising firms to adjust their ads to currents needs and general preferences by highly reducing cost for identifying such potential consumers.[9]

Furthermore, not only the exposure to certain ads, but also the impact a particular ad had on the exposed audience can be squeezed out of collected datasets. By linking online ads to the subset's later online behavior (incl. purchasing, browsing, survey responses, etc.) the clickstream can generate data for comparison of the purchasing behavior between individuals being exposed to different ads for the same product, for example. Collecting data through website linkage therefore can enable causal measures of advertisement effectiveness.[10]

Although there have been earlier models of customer data collections such as purchasing records of customer loyalty cards, telephone sales records or surveys conducted by polling firms, producers or service providers did not have the means to monitor the behavior of existing AND potential customers at the same time to the same extent as today offered by big data analysis.[11] Today, given the technical possibilities to match and split collected data through cook-

[7] Strahilevitz, p. 2012.

[8] Goldfarb/Tucker, p. 6.

[9] Goldfarb/Tucker, p. 7.

[10] Goldfarb/Tucker, p. 7.

[11] Goldfarb/Tucker, p. 10.

ies or agreements with other websites, firms can turn their business operations more effective by tailoring products to specific consumer preferences and needs, by the possibility of immediate feedback or by developing recommender systems to customers. Such systems display other purchases made by people who have one purchase in common (e.g. the purchase of a specific DVD), thus again trying to identify potential target consumers for another product.

As stated above, the collection of data only touches privacy concerns, when this data is of an identifiable nature. The collection of data regarding browse and click related information might not directly be linked to individuals. Of course, when several data sets can be linked together, seemingly impersonal data may indeed allow interference with an individual's identity. However, Internet users may find their privacy also touched in cases where other people using their personal computer can draw conclusions from the ads being shown in their search engine. As an example, advertisement on credit consolidation might lead others to think the (a, respectively) person using this search engine on that particular personal computer might have financial problems – an information hardly anybody would like to disclose to any others.[12]

Deliberate Participation in Privacy Related Data Collection

In Chapter III.2.2 Zarsky outlines the fact that users continuously use applications such as Google Street View and Amazon although they sometimes are in conflict with FIPPs. Users of Google Street View for example would not consider it a privacy infringement and refrain from using the application because their house or street they live in is visually accessible to world though the application. To some extent, the public seems to have adapted to possible (light) drawbacks in privacy and does itself not consider such as harmful use of personal information. As a matter of fact, what used to be considered as private only some years ago (friends, favorite meals or restaurants, vacation plans, etc.), has become highly public today.[13]

[12] GOLDFARB/TUCKER, p. 3.
[13] CAVOUKIAN, p. 12.

Furthermore, there are examples where consumers / users actively and deliberately participate in the collection of personal data. In her paper, Julie Cohen explains the connection between so called "gamified" environments and surveillance. Feedback based rewards motivate users to engage in continual self-monitoring which eventually ends in surveillance by the application provider. In the examples she states (Foursquare, H&M, Groupon, Nike+), personal information about the user is collected at the time of registration, as well as, continuously during the time of "play". Such information is then not only used to grant the promised rewards like discounts, but again, also for targeted marketing. Turning a purchasing-related action into a game that rewards customer loyalty with social recognition makes the user, customer respectively, more ready to disclose personal information.[14] For example, the endless competition in Twitter and also Instagramm for followers, retweets and favourites awakens a more intrinsic motivation to the sharing of personal data. The roots of such commercial surveillance environments lye – among others – in the Quantified Self movement, founded in 2007 in order to provide better living though data. Although the providers of services for health tracking, diet and fitness promised to keep the user's data safe, commercial providers entered the field which eventually lead to a shift from deemphasizing data control to emphasizing the need to continuously and collectively provide and share data to control various aspects of life, such as agendas, sleep patterns, weight, and general health data, etc. The reason why such gamification of the Quantified Self movement is appealing to users might be the fact, that it turns participation and sharing into pleasurable activities, or as Cohen states it, in the digital world "sharing is caring". What is more, the collective use of such gamified environments can reach such a highly demanded level up to the point where being a player is socially valued and "gamic death entails a form of social death".[15]

Innovation Without Compromising Privacy?

The examples stated above show that there are situations where users / consumers are willing to provide or share information where in other even similar

[14] COHEN, p. 2.

[15] COHEN, p. 3 f.

situations, they are not. Also some might feel harmed by the use of certain information, while others will not in any way. As there are extrovert and introvert, sophisticated and unsophisticated, rich and poor users, the respond to various forms of data use and possible privacy infringement is very heterogeneous. [16] In the *privacy -> trust -> innovation* perspective, Zarsky puts forward a possible strict libertarian argument, that firms already have sufficient motivation and incentives to meet requested privacy standards which as a consequence would turn privacy regulation unnecessary altogether. The most important incentives to offer high privacy standards there would be own business interests. As rumor has it, no firm would like to be involved in privacy breach scandals such as recently happened with UPC Cablecom in Switzerland, where a news broadcast uncovered security leaks in the accessibility of customer related information.[17] Such weakened customer loyalty might eventually result in revenue and profit losses.[18] However, Zarsky rejects this argument with the fact that not only individuals, but also firm managers might act irrationally, thus ignore long-term disadvantages like reputation harm in the sake of short-term (financial) gains. Firms might therefore not be motivated – or even wise enough – to provide adequate and sufficient privacy protection on their own.[19]

As outlines throughout Zarsky's paper show, neither a regulation that is too narrow, nor the possibility of an opting-out mechanism for users is in favour of the development of innovation. The problem with strict governmental regulation is the inability of the state to quickly and accordingly react to either technological or social changes. Zarsky puts forward the argument that methods and means that are considered unlawful might appear in another light even in the near future. The opting-out mechanism on the other hand, proved to fail in situations, where users did not want to provide one sort of information, however, they would have been ready to provide other, in there eyes maybe less private, information.

[16] Strahilevitz, p. 2022, 2024 ff.

[17] http://www.srf.ch/news/schweiz/grosses-sicherheitsleck-bei-cablecom, visited on October 24 2014.

[18] Cavoukian, p. 12.

[19] CTCII Final Study Report, p. 36; Tor, p. 19.

A possible answer to this either-or-policy, might be the empowering of the users themselves to control what information is used how and when. In Chapter III.2 Zarsky also outlines two of the rules of Fair Information Practise Principles (FIPPs): “notice & choice” and “secondary use / purpose specification”. As to the informed consent prior to data collection, Zarsky argues that it is difficult for firms to identify future usage of data in advance. This fact will ultimately decrease innovation capacity as firms would be distracted of innovating by trying to sort out possible “informed consent breaches” with the users.
As to “secondary use / purpose specification “, where collected data can only be used for predetermined tasks to which the user gave his ok, Zarsky states that already the proof, if such principle was breached, sometimes is very complicated. Again here, firms would have to put too much effort in finding out, weather the use of specific data would be really compatible with the purpose specification, thus again slow down the innovation process.

In 2011, Tucker examined the effect once Facebook changed its policy in 2010 and gave users increased control over their privacy settings. In her conclusion, such enhancement of privacy control in the hands of the users themselves turned personalized advertising more effective which may lead to the conclusion, that data and therefore privacy protection can be provided by more than a binary choice on weather to have protection or not.[20]

Zarsky in Chapter II.2.(4) argues against the fact, that privacy drives innovation through the fact that, when wanting to comply with either private or governmental regulation, firms have to innovate in order to generate better and more efficient solutions for their customers. In his opinion, innovations that are being brought about in such settings do not reflect the actual market need, but are the outcome of a compliance friendly steered attempt of innovation. However, this notion seems to disregard the fact, that not all information will lead to innovation or, that for information to be understood rightly, context is crucial.[21] In her paper, Ann Cavoukian argues that the users individual knowledge and consent increased the collected and analyzed data’s quality. As an example

[21] CAVOUKIAN, p. 3.; FRANK, p. 57.

for her thesis, she outlines Googles difficulties when interpreting flu-related searches as actual geographical incidence of flu-related illnesses. Due to a lack of contextual information, Google did not know why people were searching for "flu" (because they were ill themselves, because they just wanted to know the best methods not to fall ill or out of pure interest). Accordingly, by analyzing only the information on *who*, *when* and *where* was searching for Ebola-related information, Google would not be able to map the geographical incidence of Ebola due to the lack of informational data on *why* people googled Ebola. Of course, machine translation, data mining in general, as well as, machine learning are efforts to constantly improve the interpretation of provided data. This is a point that also Zarsky already stresses. However, a much easier way to improve the accuracy of data interpretation would be through the consent of the user providing it.

Another, or rather an additional approach put forward by Cavoukian are the methods of handling the challenge that lies in the risk of creating automatic linkage between apparently non-identifiable data, thus turning the whole into subjects to individual privacy.[22] Cavoukian advocates for a rather pro- than retroactive approach to privacy. Firms should therefore built in privacy protection measures already in their own technology, business strategy, and operational process.[23] This so called "Privacy by Design" follows seven principles: prevention, not remedy of data protection breaches; privacy as a default setting; privacy as an embedded element into design and architecture of IT; a positive-sum approach (win-win) to the trade offs between privacy and innovation; ensuring full lifecycle protection by securely collecting, using, retaining and destroying data; ensuring visibility and transparency of business practices and technologies by keeping such verification open to the individual user; and ensuring respect for user privacy at all times. Possible strategies to follow such principles could lie in data minimization, de-identification and – as already stated above – user access controls.

[22] CAVOUKIAN, p. 11.

[23] CAVOUKIAN, p. 14.

According to Cavoukian, data minimization has the strongest impact on managing data privacy risks as it tries to eliminate risk when data is collected, therefore at the earliest stage of the information life cycle. Data minimization on the one hand means, that data should only be collected for an antecedent defined purpose and for nothing else. On the other hand, minimizing date could also mean to simply summary or aggregate data which in most cases would already meet the firms needs to extract the relevant information.[24]

De-identification is a step that can be taken at the stage when data has been collected already. Through measures like deleting and masking direct identifiers, such as names or credit card numbers, or through suppressing and generalizing indirect identifiers, such as postal codes or birth dates, the collected data sets can be striped of all potentially identifiable information.[25]

As privacy does, contrary to security, not only relate to the way of information accession and protection, but also to the way of its collection and use, user access controls such as “informed consent” and “secondary use / purpose specification” are an effective means to diminish internal threats. When combined with other Privacy by Design policies, they can hinder possibilities of accidental or intentional disclosure or misuse of information.[26]

Conclusion

As privacy and innovation are linked to each other by a privacy -> innovation perspective, as well as, a innovation -> privacy perspective, neither the arguments for strict regulatory measures, nor a completely lenient policy with the absence of any regulatory measures on data protection will lead to instruments that are able to adapt to the ever-faster changing conditions in the ICT realm. Making companies that collect and use data solely responsible for the accurate protection of data privacy eventually also fails to meet the context-related

[24] CAVOUKIAN, p. 16.

[25] CAVOUKIAN, p. 18.

[26] CAVOUKIAN, p. 20.

needs of individuals disclosing such data. As in some cases individuals are willing to provide and share their information, yet in other cases they are not, giving users the control over the collection and use of their data by control over privacy settings through FIPPs does serve privacy protection, as well as, the development of a innovation friendly surrounding. Although such innovations may seem steered, the implementation of mutually both, governmental, as well as, private regulation will create a virtual and real universe maybe not of the highest efficiency, but definitely of high individual feeling of security which itself is worth achieving already.

References

CASTELLS MANUEL, The Rise of the Network Society: The Information Age: Economy, Society and Culture, Volume 1, New York 2009.

CAVOUKIAN ANN / STUARD DAVID /DEWITT BETH, Using Privacy by Design to Achieve Big Data Innovation Without Compromising Privacy, Information and Privacy Commissioner Ontario, Canada, 2014, available at: http://www.privacybydesign.ca/index.php/paper/using-privacy-design-achieve-big-data-innovation-without-compromising-privacy/ (visited on Oct. 24 2014).

COHEN JULIE E., The Surveillance-Innovation Complex: The Irony of the Participatory Turn, in: Berney et al (eds.), The Participatory Condition (University of Minnesota Press, 2015, forthcoming), Minnesota 2014.

Commission on Towards a Competitive European Internet Industry, Final Report, Brussels 2012, available at https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/FI3P%20Final%20Study%20Report%20v1%200.pdf (visited on October 24, 2014).

FRANK ROBERT H., The Darwin Economy: Liberty, Competition, and the Common Good, Oxfordshire 2001.

GOLDFARB AVI / TUCKER CATHERINE, Privacy and Innovation, NBER Working Paper No. w17124, Cambridge MA 2011.

POLANYI KARL, The Great Transformation: The Political and Economic Origins of Our Time. Boston 2001.

STRAHILEVITZ LIOR JACOB, Toward a Positive Theory of Privacy Law, in: Harvard Law Review, Vol. 113, No. 1 2013, pp. 2010-2042.

TOR AVISHALOM, Understanding Behavioral Antitrust, in: Texas Law Review, Vol. 92 2013.

TUCKER CATHERINE, Social Networks, Personalizes Advertising, And Privacy Controls, in: Journal of Marketing Research, Vol. 51, No. 5 2014, pp. 546-562.