REORIENTING DATAFICATION? NEW ROLES FOR USERS IN ONLINE PLATFORM MARKETS

Tuukka Lehtiniemi
Helsinki Institute for Information Technology
tuukka.lehtiniemi@hiit.fi

ABSTRACT

Personal data storage and management services aim to provide people with control to the collection, storage and use of their data. This paper investigates the new roles such initiatives propose for people in the online economy, and their effects on the markets where decisions on the uses of personal data are made. To investigate these new user roles against the existing ones, I employ Shoshana Zuboff's 'surveillance capitalism' as the value creation model of mainstream online platform markets. The identified new roles for users are data collector, intermediary of data between services, controller of data analysis, and source of subjective data. These roles work to restructure value creation from personal data, and through them the personal data storage and management services seek to shape new markets in which users are positioned as active participants seeking to benefit from their data.

1 INTRODUCTION

Online platform companies act as intermediaries of multi-sided markets: they make it possible for two or more different groups of end-users to find each other and perform exchanges (Evans 2011). The products that facilitate these markets are the internet platform services themselves: a search platform enables transactions between users, content providers and advertisers (Rieder and Sire 2013) and a social media platform makes users, advertisers and application developers meet (Helmond 2015). Online platform companies facilitate markets in order to profit from them, and there are marked similarities in how this is achieved in practice. In particular, users get the services of platforms free of charge, while businesses are on the profit-turning side of the market; and the services provided to both users and paying customers are based on the collection and leveraging of data about the users.

The businesses of online platforms are made possible by datafication, or the transformation of the social actions of their users to quantified data (Mayer-Schönberger and Cukier 2013, 78). Datafication is due to the capability of the underlying information technology not only to automate but to also informate (Zuboff 1985), i.e., to produce information on what it automates. Platform companies employ data acquired through informating to predict and modify the behavior of their users. They form predictions of the habits and interests of users (Van Dijck and Poell 2013), and also shape the context of the choices that the users do in order to channel them towards desired decisions (Yeung 2016). Fundamentally, the consequences of datafication depend on the answers to questions on what data is stored, what is learned from it, who is allowed to do the learning, and who decides about these things (Zuboff 2015). According to critics, there are significant asymmetries in the online between the capabilities of those who collect and make use of data, and of their sources and targets (Andrejevic 2014). The extensive commodification of personal data has been seen as the root cause of these imbalances (Crain 2016).

Shoshana Zuboff (2015) argues that the answers to the questions about data and what is learned from it are shaped by the underlying institutional market form of platform companies. Zuboff observes that in each era, market economy seems to gravitate towards a dominating underlying model of value creation that becomes the taken-for-granted context in which companies operate. In the era of industrialism, a dominating model was corporate capitalism based on mass production. Zuboff argues that in the online space, the dominating model is a specific form of informational capitalism (Castells 1996) pioneered by Google and shared not only by other large companies, but also by default most online startups. She calls this emerging model "surveillance capitalism": the extraction of data about people, with the aim of predicting and influencing their behavior for profit. Its underlying assumptions define what data is extracted, who participates in the production of predictions, and who can benefit from the predictions. The answers are formed, says Zuboff, in markets of behavioral prediction and modification. Importantly, these markets are not consumer-facing, and the users do not participate in them. It is in this sense that users are the products, and not the customers, of commercial surveillance (Crain 2016).

Even if surveillance capitalism is the default market model in the data economy, the market economy also spins out alternative models as a response to new technological affordances and to the needs and wants of consumers. In this paper, my interest is in initiatives that try to, as Spiekerman et al. (2015) put it, place people on the driver's seat for their personal data. They aim to make it possible for people to collect and store their personal data, and control its use by others (Abiteboul et al. 2015). As such capabilities to control data do not generally exist within platform services, these initiatives present a potential break from the surveillance model, or a reshaping of surveillance markets. The questions I present in this paper are: What are the specific roles these initiatives offer people with respect to datafication and surveillance? What are the consequent ways in which they attempt to reorient the surveillance markets?

In order to answer these questions, I first develop a lens of the current role of people with respect to datafication based on Zuboff's description of surveillance capitalism. I then employ this lens to the analysis of three personal data storage and control initiatives based on developer interviews and questionnaire responses. In section 2 I present in detail the surveillance capitalism model and the position it has for users. I describe the cases in Section 3, and in Section 4 I identify and analyze the roles they propose for their users. Section 5 concludes with discussion on how these roles work to reorient the markets of behavior prediction and modification.

2 THE MARKET FORM OF INTERNET PLATFORMS

2.1 PLATFORM COMPANIES AS MARKET INTERMEDIARIES

Tarleton Gillespie (Gillespie 2010) traces the position of online content providers in relation to users, customers and policymakers by examining the use of the term "platform". Gillespie analyses how particularly the strict computational understanding of the term, where the platform is an infrastructure that enables the development and deployment of applications, has been relaxed from its technological meaning to more loosely describe the online services of content intermediaries. Annabelle Gawer (2014) identifies two distinct theoretical perspectives to technological platforms: the engineering perspective in which platforms are viewed as modular technological architectures, and the economic perspective that is relevant to this paper. In the economic perspective, platforms are viewed as intermediaries of double-sided (Rochet and Tirole 2003) or in more general, multi-sided (Evans 2011) markets. A company operating the platform creates products or services that facilitate exchanges between different types of market participants, and by doing so, creates markets. It creates value due to the inability of the exchange participants to find each other or to perform exchanges without the platform, or due to

the reduced cost of doing so with the aid of the platform. The platform operator's pricing strategy often entails selling products on one market segment below cost (Rochet and Tirole 2003). Losses on one market segment are incurred in order to stimulate the sales of products in other, profit-turning market segments, that subsidize the loss-incurring segment.

Online platform companies, then, are businesses that set up market-intermediating technologies for the purpose of creating, and with the aim of capturing a portion of the value for themselves. Towards this end, they organize the platform interactions in a way that they can leverage for profit. Many offer services for free on the market where consumer end-users participate, and turn profits on other markets. Viewing Google search as a platform in this sense, Rieder and Sire (2013) identify three distinct parties whose interactions the platform mediates: users, content providers, and advertisers. These interactions take place on two markets. On one of them, the search service allows the users and content providers to meet. On the other market Google sells targeting to advertisers. The provision of the targeting services is based on the search market: advertisements are displayed to users beside the search results, and the display of advertisements is based on data collected and information gained from that segment. Rieder and Sire show the ways in which the platform operator has incentives to influence the users' actions on the consumer market in ways that enable maximizing revenue from advertisers on the other market. Analyses have been made also on the markets organized by other internet platform companies. In Anne Helmond's (2015) techno-economic approach, Facebook connects users, advertisers, and third-party developers. Helmond examines the relationship between the technical architecture and the economic model of the platform: Facebook extends itself into the web, commodifying web content and user activities also outside the platform in order to format this external data for its platform and to suit their economic model.

These companies seem to organize their platform-enabled markets, particularly the participants' relations within these markets, with similar principles. Selling targeting is a key source of revenue for online platform companies. The users' side of the market typically incurs losses that are subsidized on the advertisers' side. As I proceed to describe next, also a more general underlying logic exists according to which the platform companies operate.

2.2 SURVEILLANCE CAPITALISM

The outward similarities between the value creation models of large-scale internet platforms, and hence between the models of their markets, raises the question of a common underlying logic according to which these platforms operate. Shoshana Zuboff (2015) argues that the success of the market economy is based on the ability to produce both new markets and new institutional forms of markets. The institutional form reflects the assumptions underlying the models of value creation of businesses and their positioning towards market participants. Zuboff characterizes the emergent institutional market form of the online space as surveillance capitalism. Its model of value creation is to produce "objective and subjective data about individuals and their habits for the purpose of knowing, controlling, and modifying behavior to produce new varieties of commodification, monetization, and control" (ibid., 85). The multi-sided markets operated by the large-scale internet platforms are instances of this surveillance model, and its assumptions are embedded into the ways platform companies organize their markets and collect, store and use personal data about their users.

Zuboff describes the value creation process as taking place in three phases: data extraction, behavior prediction, and monetization of predictions. At the first phase, the company provides products or services for people to use, and targets the users with ubiquitous extraction processes to collect data about them. The users become the sources of what Zuboff calls *surveillance assets*, a raw material for later phases of production. At the next phase, the company uses the extracted

data as input material to produce *prediction products*[1] from surveillance assets. The conversion of surveillance assets to prediction products happens by employing highly specialized analysis capabilities, *surveillance capital*. Predictions include qualities, preferences, characteristics, intentions, needs and wants of users. The third phase is about converting prediction products into revenue. In the surveillance model, revenues come from beneficiaries of prediction products, most famously but not necessarily advertisers. These phases continue with new transactions producing new possibilities to extract raw data. For example, employing a user profile based on extracted data to target advertising can lead to clickstreams and eventually concluded economic transactions, the data about which is valuable for the purposes of further targeting. Value creation is a recurring process, with data extraction, analysis and revenue extraction taking place in an ongoing and simultaneous manner.

In what follows, I have extracted features of Zuboff's description of surveillance capitalism and its logic of value creation from personal data with the specific interested in the positioning and roles it offers to individuals. Zuboff's description is detailed and can be operationalized for analysis that allows gaining insight into roles that personal data storage and control initiatives strive to provide to individuals.

Data extraction

The process by which the surveillance capitalist collects personal data about the users is *data extraction*. It is an essentially one-way process that does not entail reciprocities or dialogues between the company and the users (Zuboff 2015, 79). The extracted personal data signals personal and potentially intimate details of the users. Formally, extraction happens with the consent of the users. It has been argued by privacy scholars that in the context of ubiquitous data collecting and advanced data analytics techniques, consent is unlikely to be voluntary or have meaningful content (Acquisti et al. 2015; Solove 2013; Yeung 2016). It is also notable that the extraction of data increasingly takes place also beyond the platforms, essentially expanding the internet platforms' capacity to extract data to, for example, the open internet (Gerlitz and Helmond 2013) or to include various devices or sensors.

Quantity of data

Due to the probabilistic nature of its analytic capabilities, the surveillance capitalist primarily values *quantity, not quality, of data* (Zuboff 2015, 79). Actions of the users, no matter how trivial, are considered signals to be analyzed and fed back (Mayer-Schönberger and Cukier 2013, 113). The surveillance capitalist benefits from more data as it improves the predictive value of analyses, leading to increased possibilities to turn them into revenue. The company has incentives to collect as much information about the users as possible (Rieder and Sire 2013). This is true not only regarding the breadth of data about a given user, but for the breadth of the user base as well. In addition, when extraction and analysis of data about user behavior improves service quality, extracting more data leads to more users and advertisers choosing the particular service, which again leads to better service (ibid.). According to the surveillance logic, accurate and validated data points about an individual user are not as interesting for the provision of predictive analyses as broad data from many sources.

Accumulation of rights

Zuboff conceptualizes privacy as the capacity of making a choice on the spectrum between secrecy and transparency (2015, 83). The right to privacy, then, is a decision right concerning the preferred position on this spectrum. In addition to accumulating extracted data, the platform

[1] Zuboff introduces this term in a later essay (Zuboff 2016).

company also *accumulates decision rights* concerning the data they extract by means of asking consent of the user. Consent is provided in an environment characterized by lack of transparency to data practices, context-dependent and malleable attitudes towards privacy, un- or misinformed decisions regarding disclosure of data, and various difficulties of making considered disclosure decisions in the first place (Acquisti et al. 2015; Solove 2013). In practical terms, the means for providing a meaningful consent are limited (Zuboff 2015), and therefore the decision rights are redistributed from users to platform companies.

Production of predictions

Not only the extraction of data, but also the *production of prediction products* based on the extracted data takes place without feedback loops to the user (Zuboff 2015, 80). The prediction products end up signaling things about the users, but the users themselves may not be even aware about their existence. The predicted things – characteristics, preferences, traits, details – are potentially intimate and may turn out to be consequential, but the users have limited possibilities to view, accept, deny or correct these predictions. In addition, they have limited access to information needed to comprehend the process that leads to these predictions.

Asymmetric positions

The position of the surveillance capitalist and its users are decidedly *asymmetric* due to the differences of access to capabilities of data collection and analysis (Zuboff 2015, 83). The company operates via employing specialized means of production that rely on proprietary knowledge and material capabilities. The users do not have inherent capabilities to draw inferences on their data, nor do they have access to the analysis capabilities employed by the company. In addition, the proprietary capabilities include data assets that are potentially extracted over a wide range of users and data sources. In other words, the production capabilities of the company rely on its position as the aggregator of data about many individuals and from multiple sources. The material and knowledge asymmetries institutionalize the lack of reciprocities between the company and its users (Zuboff 2015), both in the data extraction phase and the analysis phase, an asymmetry described as the "big data divide" by Mark Andrejevic (2014).

Power asymmetries

These asymmetric capabilities also give rise to *power asymmetries*. Power in surveillance capitalism arises through the control of means of behavior prediction and modification (Zuboff 2015, 82). This includes controlling opportunities to employ ubiquitous data extraction and analysis capabilities. More specifically, through this control, the surveillance capitalist is able to exercise calculative power (Callon and Muniesa 2005) over the market participants of the platform-enabled markets. This means it is capable of assessing the value of the data it extracts of the users, and at the same time able to affect the users' possibilities to perform the same valuation. By exercising this power, it can essentially prevent the users from economic action towards their personal data. Power asymmetries, then, have repercussions regarding the position the users can assume in the platform-enabled markets.

Source of revenue

In the end, the *source of revenue* for the surveillance capitalist is the sales of prediction products. However, the access to the means of behavior prediction and modification are determined on a market that the users do not participate in (Zuboff 2015, 85). The products that turn profit for the surveillance capitalist – such as advertisements – are not something that the users could even feasibly purchase. This means that the consumers, who are the users of the company's services,

are not its paying customers. Micro-level analysis of, for example, Google's tangled activities on the different segments of the multi-sided market shows it has incentives to organize its interactions with the users in a self-serving way (Rieder and Sire 2013) in order to optimize its revenues.

Formal indifference

To summarize, the three phases of value creation process – data collection, data analysis for behavior prediction, and monetization – are characterized by a lack of feedback loops or reciprocities to the users. The surveillance capitalist unilaterally controls the conditions of production of prediction products from raw data (cf. Crain 2016). An inherent feature of the market logic described by Zuboff is the *formal indifference* of the company from its users. The company is indifferent to the contents of data as such. It does not care what the users say and do, as long as they say or do it in ways that can be turned into data assets (2015, 79). The other features listed above – asymmetric positions of market actors, the lack of reciprocities and feedback loops between the company and its users in different phases of production, data extraction practices and motivation, power aspects and roles in the marketplace – all indicate or work towards institutionalizing this indifference.

3 INITIATIVES TO REORIENT DATAFICATION

3.1 BACKGROUND AND APPROACH

The idea that there could be an intermediary to facilitate personal data interactions is not exactly new. Initiatives that aim to provide people with control over their personal data seem to be, to an extent, reiterations of the idea of an "infomediary", or an information intermediary, that dates back to late 1990s (Hagel and Rayport 1997). People would claim ownership of their personal data and would be willing to make data available if they stood to benefit from it. People would then become information suppliers, and a third party, the infomediary, would be needed to facilitate information transactions by connecting information supply and information demand. The infomediary, then, would facilitate markets of personal data through which people would receive beneficial services in exchange for their data. With hindsight, the envisioned infomediaries did not emerge in the late 1990's, and online platform companies are currently providing many of the services envisioned at the time and vastly more without infomediaries. Instead, these services are produced based on accumulated data within the large-scale internet platforms.

In recent years, a renewed interest has emerged towards services for storing personal data and controlling its uses (Abiteboul et al. 2015). Several examples of such services and concepts have emerged from commercial developers[2], academia[3] and non-profits[4]. A market report on "personal information management systems" identified some 30 such systems in different domains in 2014 (Ctrl-Shift 2014). Interestingly, large players in the online economy seem to be exploring similar concepts (Gurevich et al. 2016). The specific approaches include sector-specific services e.g. for health data, sector-agnostic personal data storage services, and standard-like data interoperability frameworks. While there are marked differences between these approaches

[2] Such as Cozy Cloud (https://cozy.io/en/), digi.me (https://digi.me), Healthbank (https://www.healthbank.coop), Meeco (https://meeco.me), and Qiy Foundation (https://www.qiyfoundation.org)

[3] Such as Databox (Chaudhry et al. 2015), DiMe (http://hiit.github.io/dime-server/), Hub of All Things, (http://hubofallthings.com/), and OpenPDS (de Montjoye et al. 2014)

[4] Including the health data cooperative MIDATA.coop, (https://midata.coop)

and it is impossible to say anything about their eventual success, the underlying vision of how things should be in the digital world seem to have some common elements: people should have more control towards their data, and this would lead to benefits for both the individuals themselves and for service providers that wish to use their data. One way to spell out this vision is that currently it is organizations that are in control of personal data about their users or customers, and this should be made more human-centric instead (Poikola et al. 2015).

Following the research questions posed in the beginning of this paper, I set to examine the potential reshaping of the surveillance markets that these initiatives present by looking at the specific roles they offer people with respect to datafication and surveillance, and the consequent ways in which they attempt to reorient the surveillance markets. Towards this end, three personal data storage initiatives that aim to carve themselves a position somewhere between individuals and data-using companies were selected for closer scrutiny. Two of the case initiatives are startup companies (Meeco and Cozy Cloud), and the third is an example of the stream of research-originated technology innovations from the MIT Media Lab (OpenPDS). The motivation for selecting these cases was that they exhibit both similarities and contrasting qualities between them. All three aim to enable the individual to store personal data into a personal repository, and to make use of this data by providing it to purposes that they deem beneficial. They then exhibit the general aims of storage and control of data (Abiteboul et al. 2015). Outward differences include the origin of the developer from France, Australia and the US, and the consequent potentially different cultural orientations towards e.g. privacy. The type of data that the cases would have people control spans from all sorts of mundane, everyday data to log-type, behavioral data. Their stated aims range from overarching management of digital life to privacy-conscientious provision of data in exchange for services. The developer's status as either a startup company or an academic research project affects expectations of the underlying immediate economic motivation. Finally, their expected target audiences vary from advanced to general users.

As these cases represent a potentially emerging field or industry, their representativeness of the field, maturity, user base or success potential with respect to any other initiatives are open questions. I view the work such initiatives do as dynamic and ongoing market shaping, without clearly successful leaders or "killer apps" at this moment. Nevertheless, exploring the field by concentrating on a small number of cases allows us to identify their market-shaping aims at a detailed level. A detailed analysis of shaping new markets also opens up a view to how markets work currently. Analyzing a dynamic situation, of course, runs the risk of further dynamics that may undermine the analysis. Despite this, I consider this approach relevant. Initiatives resembling the cases seem to be widely emerging at the moment, and it is important to understand what they are and do. In addition, their emergence is supported by regulatory developments such as the adoption of the new EU General Data Protection Regulation (GDPR) (EU 2016) and its rules on portability of data in machine-readable format.

The empirical data on the three cases consists of interviews with a member of the developer team, responses to an open-ended questionnaire, and publicly available materials. The interviews were explorative in nature and done with the general goal of gaining understanding of what these actors aim to achieve and why. The open-ended questionnaire was conducted by the European Commission for the purpose of gathering background information for a roundtable of "personal information management architectures" developers.

Interviews and questionnaire responses were first analyzed based on open coding. At later stages the analysis focused on sections that concerned the positions of the end-users of the services. Additional insight was sought by comparing and contrasting the interview findings to mission statements and feature descriptions available via public sources, including the websites of the

initiatives and, where available, research publications. Where possible, also first-hand use experience of their product offerings was sought with the same aim. Based on this analysis, the functionalities and features that the cases aim to provide their users were distilled into four user roles. The overall aims and features of the cases are described next. After presenting the cases, I proceed to describe the new roles for users they propose.

3.2 CASE DESCRIPTIONS

3.2.1 COZY CLOUD

By their own definition, Cozy Cloud[5] is a personal private cloud. Its intended use is to store personal data within, and a to install applications that make use of the data. The kinds of data that are to be stored include various data that are otherwise spread and siloed in different clouds, services or platforms: for example, mundane everyday data such as photos, emails and documents, banking or other financial data, health data, or data produced by fitness trackers or Internet of Things appliances. The intent of Cozy Cloud is to provide a place where third-party applications can, with the permission of the user, make use of personal data without the need to send it anywhere; the data and the application would remain in the personal cloud.

In practical terms, Cozy Cloud is open source software that a user may install, run and administer on their own machine or on a virtual server. Setting up Cozy Cloud this way requires technical skills, and thereby it is oriented towards somewhat skilled users. Alternatively, it may be provided as a paid or add-on service by, e.g., a hosting service provider. In any case, each user has their own instance of Cozy Cloud running on a private or virtual server.

The user, then, accumulates data from various sources to a private space, and services that make use of personal data, such as data analytics, are run in this space. Cozy Cloud's model proposes four conceptual changes in comparison to the surveillance model. They relate to who accumulates personal data, who can make use of it, where analytics and data use happens, and who profits from data. Instead of accumulating data to proprietary servers owned by service providers and platform companies, data is accumulated to the user's server. Instead of only the primary data collector and downstream parties that get hold of collected data, any service providers can employ data in the personal cloud as long as the user provides access to it. Instead of data and analytics residing in a service provider's server, they reside in the user's server. Finally, instead of profiting from the users' personal data, Cozy Cloud as the platform provider aims to profit from application development and from the provision of the cloud service to business customers. The users, however, may agree to allow the use of their data for any purposes by third party service providers, including marketing.

3.2.2 MEECO

By their own definition, the purpose of Meeco is to "help you manage life and all your important digital relationships"[6]. According to Meeco, the current practices of advertising and targeting and the related data brokering, online tracking and data analysis are not efficient and lead to low quality data about people. Data acquired directly from people could, in their view, be more accurate and contextually and personally relevant. A direct exchange of this kind of data between individuals and businesses or other users of data would then lead to more value for both sides than the current industry practices. In this way, they intend to make it possible for people to exchange their personal data to things they value, such as personalization, offers or insights. The thinking underlying the Meeco model draws on the idea of the intention economy (Searls 2012),

[5] https://cozy.io/en/ (accessed 31.8.2016)
[6] https://meeco.me (accessed 31.8.2016)

which refers to a market in which buyers notify the markets of their plans, needs or purchasing intentions, instead of sellers attempting to guess these.

Towards this end, Meeco provides a platform service for users to store and share data about themselves. In practical terms, Meeco is a cloud storage service in which users create accounts. Within the service, users arrange their data in datasets associated to specific things such as other people, items, places, concepts, and intentions. Contents of these dataset could include, e.g., attributes or characteristics, preferences, measurement results, or connections to other things. Users can then share some of the contents of these datasets with businesses or other actors, for a specified purpose and time, in exchange for benefits expected to be derived from the data.

In comparison to the surveillance model, Meeco also promotes changes to who accumulates data, and who can make use of it. Instead of the platform provider, the users themselves accumulate data, and for the purpose of sharing it to various service providers. In addition, Meeco specifically proposes to make individuals the primary sources of things like habits, preferences, or intentions. With this, it aims to circumvent stages of production of prediction products based on extracted data: instead, users directly provide accurate data in exchange for benefits. Finally, Meeco states clearly that data stored in the service belongs to the user, and is not mined or used by the cloud provider for profit.

3.2.3 OpenPDS

OpenPDS[7] is a personal data storage service based on research carried out at MIT (de Montjoye et al. 2014). Its developers argue that people do not get the best possible services that could be provided with their data due to difficulties in providing access to data and in privacy preservation. Its developers envision that when users control access to data via openPDS, they are provided with rights to possess, have full control on the use, and to dispose the data (Pentland 2009). With these rights, users are expected to be able to choose the best algorithms for their data based on whether a service provides enough value for them taking into account the amount of data it asks to have. OpenPDS focuses on log-type behavioral data: data that is automatically generated by e.g. smartphones, sensors, credit cards. Difficulties of privacy preservation are connected to providing access to raw behavioral data in a way that prevents further use of data and re-identification of anonymized data.

OpenPDS is an open-source personal data storage where behavioral data about the user is stored. When the user gives an application or a service provider access to behavioral data, openPDS does not hand over the data directly. Instead, the request for data is sent to openPDS in the form of a question. OpenPDS then runs this question against personal data and sends back the answer. The aim is that processing of sensitive data happens within openPDS, and only the results are sent outside the service. People then provide services and applications with access to not the data itself, but on features that openPDS derives from the data based on requests.

Like the other cases, openPDS also promotes changes to who accumulates data, and who can use it. In addition, compared to the surveillance model, the production of predictions is not anymore based on the analysis of raw data. Instead, the data storage service performs some steps of data analysis on raw data, and only the resulting intermediate products reach later stages of production. By using the service, the users themselves conceptually perform some stages of production on raw data. In addition, by providing intermediate products instead of raw data, the users are expected to have some power to limit the possibilities of further production on their data.

[7] http://openpds.media.mit.edu (accessed 31.8.2016)

4 NEW ROLES FOR USERS

The key proposed roles for individuals that were distilled from the analysis of cases are data collector, intermediary of data between services, controller of data analysis, and source of subjective data. In this section, I will mostly refrain from referring to the singular features of each case, and instead will concentrate instead on the user roles that these features support. These user role themes can be to an extent found in all three initiatives, albeit with varying prominence and focus. This reflects the dynamic nature of the work that these initiatives perform in seeking a workable approach. What I consider important here are the roles themselves, rather than their prominence or success in a specific case.

4.1 INDIVIDUAL AS DATA COLLECTOR

One role proposed to users is to act as a data collector by using technologies to accumulate personal data in a private data repository. Several types of sources of accumulated data are envisaged, e.g., data uploaded or input by the individuals themselves, data collected by sensors or devices, or data initially collected by other online services. The latter includes data that is collected and proprietarily stored by large-scale platform companies. An individual's ability to do this rests on the capability to access and transfer personal data in machine-readable format, supported by e.g. the right to data portability in EU GDPR (EU 2016).

When users accumulate data in a private data repository, they become participants in the process of data collection. In contrast with the data extraction model of surveillance capitalism, accumulated data can become the subject of reciprocity and decisions of inclusion, exclusion and moderation by the individual: what sources of data are to be included or left out, what individual pieces of data are desired or unwanted? Contents of data within a personal repository can be also accessible to users later on, resulting in a further feedback loop with possibilities to make decisions of inclusion and removal also afterwards. Data collection and storage, then, become ongoing negotiations that can at least potentially take into account past decisions.

These features work towards rebalancing information and capability asymmetries between individuals and data collectors. The reciprocities and feedback loops are, however, obviously limited to the contents of the private repository. Data initially collected by, e.g., other service providers becomes subject to reciprocity only when it is ported into the private repository.

4.2 INDIVIDUAL AS DATA INTERMEDIARY

In the data collector role, users are able to accumulate data over various sources into their private repository. The intermediary role comes into play when users can choose to provide access to accumulated data to third parties. In this way, the individual intermediates data between data collectors and third parties. If data from proprietary silos has also been stored in the private repository, users are envisaged to be able to intermediate also data that is currently inaccessible to third-party service providers.

In the surveillance model, the production and ultimately monetization of predictions begins with extraction of data. When an individual takes the data intermediary role, the first part of this value chain is transformed: the data extraction phase is replaced by a phase where an individual is asked to provide access to data within the repository. The individual is envisioned to act as a gatekeeper between data collectors and third parties, allowing access to data if it is associated with sufficient benefits.

An important element of the data intermediation role are the terms under which data is intermediated. Various visions for user-controlled terms include temporal limits to sharing data, limiting usage purposes, moderating data after providing access to it, and revoking access. In this

way, the decision rights on personal data are to remain with the user and not to be accumulated with an external actor as in the surveillance model.

4.3 Individual in control of analytics

Down the line, the next proposed role for individuals is controlling analytics run on the stored data. There are two ways for individuals to gain access to analytics capabilities. One way is embedded in the capability to share data: users are expected to share data with service providers that can provide them with useful analytics-based services. The other way is to run analytics applications within the data storage itself, in which case the data is not necessarily shared with anyone.

With these means, users are to be put in control of producing prediction products: choosing what part of their data is to be used, what analytics are run, and to what purposes the predictions are produced. In addition, the aim to control extends to preventing third parties from using data to learn things that are not desired. Performing analytics within the private repository prevents its flow to non-intended uses. In situations when data is to be sent outside the personal repository, pre-processing raw data before providing access to it works towards the same end. The purpose of pre-processing is that not the raw data itself, but instead an aggregated, anonymized or summarized data, reaches the value creation processes of external data processors. Such control further emphasizes the intention to keep decision rights concerning data with the user.

In comparison to the surveillance model, by controlling analytics the user is to be made a participant in the process of producing predictions. This is to be made possible directly by means of providing individuals themselves with capabilities of data analysis, or indirectly by means of giving individuals the choice to which purposes and under what conditions they want to submit their data to third-party analyses. The individual either performs production by themselves, initiates the production process by choosing whom to give access to data, or performs some stages of the production process before handing out data to later stages.

4.4 Individual as source of subjective data

The fourth role proposed to individuals is to act as primary sources of subjectively relevant data about themselves. In addition to the capability of storing and sharing raw or pre-processed data, the users are provided with the capability of sharing data about their preferences, characteristics, intentions, and relationships. The assumption is that the individuals are willing to provide these accurate pieces of data when they can benefit from sharing them.

Importantly, these are the kinds of data that prediction products of the surveillance economy are typically about. Raw data by itself is not the point of extracting data about individuals: the point is to arrive at subjectively relevant predictions about individuals. The whole aim of the production chain starting from data extraction is to end up with such prediction. In the data source role, individuals themselves share data that can fulfill the same purpose as these predictions. In other words, the need for predictions is to be circumvented by enabling the individuals themselves to share subjectively relevant pieces of data.

One goal, then, is to provide technologies that allow the individuals themselves to explicitly signal things that the surveillance capitalist would otherwise try to predict. This specifically aims at increasing the quality of data that things like targeting, recommendations or personalization are based on.

4.5 SUMMARY OF USER ROLES

The features of the cases work towards reshaping the role of individuals compared to their role in the surveillance economy. This reshaping is to happen with different entry points into the value creation process: users are expected to participate in data collection, to intermediate data between services, to control data analysis to produce predictions, and to altogether circumvent the production processes embedded into surveillance economy. On the whole, by these means, the control of personal data is expected to turn into control of the means of behavior prediction.

Above, the issue of how individuals are to convert the predictions to value was also briefly touched. The roles of data collector, data intermediary, analytics controller, and data source all work to make individuals active participants in value exchanges concerning their data. The underlying assumption is that people are willing to provide their personal data for purposes from which they stand to gain something. In these proposed roles, individuals would make choices on what data they wish to store, who is given access to that data, on what terms the data is shared, and what kinds of analytics are to be performed on their data. In order to become the beneficiaries of prediction products, they are envisaged to make these choices in a way that ends up working to their own advantage.

So far, I have mainly concentrated on the roles provided for users. These new user roles also affect the positions of other actors. In surveillance capitalism, one element of market success is the capability to accumulate a variety of personal data from various sources and contexts for proprietary use, with the aim of increasing the effectiveness and value of predictions by increasing the scope of data collection. Actors that do not have access to these proprietary data assets are not able to compete with their more successful counterparts. A consequence of the new user roles is that the users would become the sources and gatekeepers of their data aggregated over various contexts. A new path to success would be to provide users with personally most relevant analytics in order to get their attention and gain access to their data. To turn this around, the benefits that would accrue to users depend on the market's ability to provide competing analytics to choose from as soon as data becomes available.

The technologies that enable controlling data are expected to balance asymmetries between companies and their populations by affecting the divisions of information, capabilities, and power in the online space. I conclude with remarks on how this relates to datafication and other ongoing movements that aim to reshape it.

5 CONCLUSIONS AND DISCUSSION

With this understanding of the roles that the initiatives hope to establish for their users, I return to the questions about datafication, the surveillance economy, and people's role in them. Six observations can now be made based on the above analysis.

First, these initiatives work firmly within datafication. In other words, they take the quantification of the everyday and the uses of data about people as starting points. The problem they target is not that personal data is commodified: rather, it is that the surveillance economy is not organized as it should be, and consequently people do not get to reap all possible benefits of commodification of their data. To solve this problem, they aim to create a new space for people within datafication.

Second, this new space is to be created by shaping new markets within datafication. In surveillance capitalism, the markets for prediction products face advertisers and other companies. In contrast, the new markets shaped by the initiatives are to be consumer-facing. The initiatives aim to reorient the markets that settle who can benefit from datafication, who gets to

decide who can benefit, and what are the terms under which this happens. People, the sources of the data, are to be made data-supplying and benefit-demanding participants in these markets.

Third, the personal data storage and management initiatives work with the assumption that new opportunities for beneficial services are opened up when there is a channel through which people can make their personal data available. As this supply of new data becomes available, it is assumed that there are service and analytics providers that are ready to make use of this data. Consequently, markets are assumed to provide new and better services for people to choose from.

Fourth, reorienting surveillance markets to face people is in practice sought by providing people with technological means to act in these markets. The new roles created with these technologies represent a particular kind of position for the user: the position of a rational individual who bases data sharing decisions on cost-benefit analysis. Currently, the ability of individuals to effectively perform such analyses is severely limited by cognitive and structural problems (Solove 2013). While the rational choice-maker is the target role, technologies to control data do not by themselves assure people are able to overcome the problems hindering cost-benefit analysis. Making it possible to choose between a larger array of analytics does not guarantee an ability to make meaningful analysis of the consequences of this choice. To effectively enable economic action on personal data, the personal data control technologies would need to be able to provide calculative tools for individuals to overcome both cognitive and structural problems.

Fifth, it is not only that these technologies work to construct particular kinds of positions for users; they are also developed with a particular kind of understanding of the user. The proposed roles assume an interested and involved individual, ready to take part in maintaining the storage of their data, and controlling its uses. The new user roles work to further individualize the managing of personal data. At the same time, these roles also come with increased responsibility that comes with the new possibilities to manage and control data.

Sixth, the extent to which the new user roles work towards enabling control of behavior modification, in addition to behavior prediction, remains an open question. The concept of the "hypernudge" (Yeung 2016) highlights the soft power approach of big data techniques to modifying behavior. With nudges, the behavior of individuals is altered in predictable ways by means of directing them towards preferred choices, but without forcefully limiting the choices that are available. Even if the initiatives are able to provide more or better control to personal data for individuals, it does not mean they can circumvent this type of behavior modification.

Finally, I relate the above to other initiatives that aim for changes to the workings of the surveillance economy: transparency initiatives (Crain 2016) and the open data movement (Baack 2015).

Examining the data broker industry, Crain (2016) observes that one theme of consumer empowerment in this industry is to increase transparency in where data about them comes from, and where it is moving. Crain argues that the goal of increasing transparency runs onto structural constraints arising from the political economy of commercial surveillance. Like surveillance capitalists in general, data brokers do not operate in consumer-facing markets. The industry is incompatible with transparency: its information sources and analytic processes are trade secrets, its information buyers and sellers are separated from information sources by complex market arrangements that defy meaningful transparency, and much of the information the industry handles is computationally generated and therefore does not have an empirical source. Transparency initiatives tend to fail in empowering consumers because they leave the underlying power imbalances intact. Crain identifies the initial commodification of personal data as the root cause of these power imbalances and suggests that due to the limits of empowerment

achievable through increased transparency, activists and policymakers should look at alternative infrastructures that could counter commodification of personal data.

What can this tell us about the potential success of the personal data storage and control initiatives in reforming the surveillance economy? A part of what they are aiming to do is certainly transparency, but consumer empowerment in these initiatives happens mainly through provision of new roles and economic positions. Proponents of the new user roles identified in this paper have a fundamentally different approach than the one proposed by Crain (2016): instead of countering commodification, they would have people as beneficiaries of commodification, and participants in the markets where these benefits are traded. Like the transparency initiatives (Crain 2016), they also posit surveillance itself as not being up for negotiation, but they aim at changing who gets to benefit from surveillance, and who gets to decide who benefits.

Baack (2015) sees the open data movement as a response to the distribution of power and knowledge due to datafication. The distribution tipped towards companies and governments is seen to impede public agency, and open data activists develop new rationalities around datafication. The activists regard raw data as a prerequisite of generating knowledge, and therefore sharing raw data is seen as a means to break the interpretative monopolies by allowing everyone to make their own interpretations. The activists, however, acknowledge that raw data alone is not enough, and both a cultural change within institutions and intermediaries that act between people and these institutions are needed to make these interpretations possible. By pushing these ideas forward, the activists aim to turn datafication to support citizens' acting in an agentic manner.

The work that the cases perform bears resemblance to the work of open data movement activists. Similarly to these activists, the cases work to reorient datafication in the favor of people, and in some sense to break the monopoly that institutional data collectors have on the data they hold. The benefits envisaged for people are to a large extent also seen to arise from sharing of data. And like the open data activists, the cases aim to create intermediaries that make these benefits possible. The kind of sharing that the cases aim, however, is not of the unlimited, free for everybody to build interpretations, and open kind that the open data activists aim at. Instead, it is to be decided on by the individual, as limited as necessary, and only done when it is individually beneficial.

The models that aim to put people in control of their data are currently in the margins of the data economy, the surveillance model of value creation being in the mainstream. Will the surveillance model remain successful in the long run? The personal data storage and control initiatives examined in this paper are not facing only an existing market they would need to reorient in order to be successful; they also face the need to reorient an institutionalized and default model of the market. The success of this attempt depends not only on the potential benefits their market model seeks to provide people, but also on the social evolution of attitudes towards commercial surveillance. Key determinants of success, then, will be whether the resigned cynicism and rationalization of the current positions (Zuboff 2015) and the feelings of powerlessness to contest the current data practices (Andrejevic 2014) can be turned into a strong enough social demand for alternative models.

REFERENCES

Abiteboul, Serge, Benjamin André, and Daniel Kaplan. 2015. "Managing Your Digital Life." *Communications of the ACM* 58(5):32–35. Retrieved (http://dl.acm.org/ft_gateway.cfm?id=2670528&type=html).

Acquisti, Alessandro, Laura Brandimarte, and George Loewenstein. 2015. "Privacy and Human Behavior in the Age of Information." *Science* 347(6221):1–4.

Andrejevic, Mark. 2014. "The Big Data Divide." *International Journal of Communication* 8:1673–89.

Baack, Stefan. 2015. "Datafication and Empowerment : How the Open Data Movement Re-Articulates Notions of Democracy, Participation, and Journalism." *Big Data & Society* 2(December):1–11.

Callon, Michel and Fabian Muniesa. 2005. "Peripheral Vision: Economic Markets as Calculative Collective Devices." *Organization Studies* 26(8):1229–50. Retrieved November 21, 2014 (http://oss.sagepub.com/cgi/doi/10.1177/0170840605056393).

Castells, Manuel. 1996. *The Rise of the Network Society. Vol. 1 of The Information Age: Economy, Society, and Culture*.

Chaudhry, Amir et al. 2015. "Personal Data: Thinking Inside the Box." Pp. 29–32 in *5th Decennial Aarhus Conference on Critical Alternatives. August 17-21*.

Crain, M. 2016. "The Limits of Transparency: Data Brokers and Commodification." *New Media & Society*. Retrieved (http://nms.sagepub.com/cgi/doi/10.1177/1461444816657096).

Ctrl-Shift. 2014. *Personal Information Management Services: An Analysis of an Emerging Market. Understanding the Impacts on UK Businesses and the Economy*.

Van Dijck, José and Thomas Poell. 2013. "Understanding Social Media Logic." *Media and Communication* 1(1):2.

EU. 2016. "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data." *Official Journal of the European Union* L119/1.

Evans, David S. 2011. *Platform Economics: Essays on Multi-Sided Businesses*. Competition Policy International.

Gawer, Annabelle. 2014. "Bridging Differing Perspectives on Technological Platforms: Toward an Integrative Framework." *Research Policy* 43(7):1239–49. Retrieved (http://dx.doi.org/10.1016/j.respol.2014.03.006).

Gerlitz, C. and A. Helmond. 2013. "The like Economy: Social Buttons and the Data-Intensive Web." *New Media & Society* 15(8):1348–65. Retrieved (http://nms.sagepub.com/cgi/doi/10.1177/1461444812472322).

Gillespie, Tarleton. 2010. "The Politics of 'Platforms.'" *New Media & Society* 12(3):347–64.

Gurevich, Yuri, Efim Hudis, and Jeannette M. Wing. 2016. "Inverse Privacy." *Communications of the ACM* 59(7):38–42. Retrieved (http://research.microsoft.com/apps/pubs/default.aspx?id=245268).

Hagel, John and Jeffrey Rayport. 1997. "The Coming Battle for Consumer Information." *Harvard Business Review* 75:53–65.

Helmond, A. 2015. "The Platformization of the Web: Making Web Data Platform Ready." *Social Media + Society* 1(2). Retrieved (http://sms.sagepub.com/lookup/doi/10.1177/2056305115603080).

Mayer-Schönberger, Viktor and Kenneth Cukier. 2013. *Big Data. A Revolution That Will Transform How We Live, Work, and Think*. London: John Murray.

de Montjoye, Yves-Alexandre, Erez Shmueli, Samuel S. Wang, and Alex Sandy Pentland. 2014. "openPDS: Protecting the Privacy of Metadata through SafeAnswers." *PloS one* 9(7). Retrieved September 29, 2014 (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4090126&tool=pmcentrez&rendertype=abstract).

Pentland, Alex. 2009. "Reality Mining of Mobile Communications: Toward a New Deal on Data." Pp. 75–80 in *The Global Information Technology Report 2008–2009. Mobility in a Networked World*, edited by S. Dutta and I. Mia. World Economic Forum.

Poikola, Antti, Kai Kuikkaniemi, and Harri Honko. 2015. *MyData – A Nordic Model for Human-Centered Personal Data Management and Processing*. Finnish Ministry of Transport and Communications. Retrieved (http://urn.fi/URN:ISBN:978-952-243-455-5).

Rieder, B. and G. Sire. 2013. "Conflicts of Interest and Incentives to Bias: A Microeconomic Critique of Google's Tangled Position on the Web." *New Media & Society* 16(2):195–211. Retrieved (http://nms.sagepub.com/cgi/doi/10.1177/1461444813481195).

Rochet, Jean-Charles and Jean Tirole. 2003. "Platform Competition in Two-Sided Markets." *Journal of the European Economic Association* 1(4):990–1029. Retrieved (http://www.jstor.org/stable/40005175).

Searls, Doc. 2012. *The Intention Economy: When Customers Take Charge*.

Solove, Daniel J. 2013. "Privacy Self-Management and the Consent Dilemma." *Harvard Law Review* 126(7):1880–1903.

Spiekermann, Sarah, Alessandro Acquisti, Rainer Böhme, and Kai-Lung Hui. 2015. "The Challenges of Personal Data Markets and Privacy." *Electronic Markets* 25(2):161–67. Retrieved (http://link.springer.com/10.1007/s12525-015-0191-0).

Yeung, Karen. 2016. "'Hypernudge': Big Data as a Mode of Regulation by Design." *Information, Communication & Society* (May):1–19. Retrieved (http://www.tandfonline.com/doi/full/10.1080/1369118X.2016.1186713).

Zuboff, Shoshana. 1985. "Automate / Informate : The Two Faces of Intelligent Technology." *Organizational Dynamics* 14(2).

Zuboff, Shoshana. 2015. "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization." *Journal of Information Technology* 30:75–89. Retrieved (http://papers.ssrn.com/abstract=2594754).

Zuboff, Shoshana. 2016. "Google as a Fortune Teller. The Secrets of Surveillance Capitalism." *Frankfurter Allgemenine*. Retrieved (http://www.faz.net/-gsf-8eaf4).