Published in New Media & Society (2017), 19*(4), 579–596*

Big Data Analytics and the Limits of Privacy Self-Management

Lemi Baruh

Department of Media and Visual Arts

Koç University

Rumelifeneri Yolu, Sarıyer

Istanbul 34450, Turkey

lbaruh@ku.edu.tr

Mihaela Popescu,

Department of Communication Studies,

California State University,

5500 University Parkway,

San Bernardino, CA 92407 USA

popescum@csusb.edu

Abstract

This paper looks at how the logic of big data analytics, which promotes an aura of unchallenged objectivity to the algorithmic analysis of quantitative data, preempts individuals' ability to self-define and closes off any opportunity for those inferences to be challenged or resisted. We argue that the predominant privacy protection regimes based on the privacy self-management framework of "notice and choice" not only fail to protect individual privacy, but also underplay privacy as a collective good. To illustrate this claim, we discuss how two possible individual strategies—withdrawal from the market (avoidance) and complete reliance on market-provided privacy protections (assimilation)—may result in less privacy options available to the society at large. We conclude by discussing how acknowledging the collective dimension of privacy could provide more meaningful alternatives for privacy protection.

Keywords: Privacy, Notice and Choice, Big Data, Data Mining, Surveillance

Big data analytics and the limits of privacy self-management

The term "big data" has become one of the most hyped marketing rally points. Estimates value big data market at $50.1 billion in 2015 (Wikibon, 2014). The rapid expansion of the big data market is not surprising: a sweep of new publications (e.g., Mayer-Schönberger and Cukier, 2013) attests to the multiple uses of big data in areas as different as digital humanities and election predictions. Closer to home, the application of big data problem-solving to a variety of commercial areas generates considerable change in how industries with direct impact over people's life chances, such as insurance, healthcare or banking, will operate in the future. For example, the use of big data in the automotive industry prompted by the advent of sensors, GPS systems, and wireless communication has restructured auto insurance models from risk-based models to habit-based models that rely on real-time monitoring to estimate personalized risk levels. Pay-how-you-drive auto insurance programs, such as Progressive's Snapshot, depend on harvesting data about real-time driving habits via proprietary data collection devices plugged into a car's own telematics systems.

That big data analytics offer potential for as-yet-understudied harm to users has also become clear. The availability of massive consumer databases has produced a thriving industry of unregulated consumer scores (Dixon and Gellman, 2014). These e-scores, developed with the help of predictive algorithms drawing inputs from increasingly massive, cross-context databases, sort individuals into desirable segments in areas as diverse as employment, tenancy, or retail, all the while keeping secret the data and method that enter into these assessments (Pasquale, 2015). The cumulative effect of missed marketing deals may eventually amount to "a mountain of pathways not offered,

not seen and not known about" (Singer 2012, online), putting a new spin on the idea of cumulative disadvantage (Gandy, 2009). Therefore, beyond the technical aspect of big data processing and its practical applications, big data seems to be generating a new social organization of knowledge that normalizes a climate of privacy loss while reproducing or even accentuating existing inequalities (Andrejevic, 2013).

Recent FTC reports (e.g., 2014) make it clear that the current regulatory environment, which emphasizes market solutions via industry self-regulation, renders the question of resistance increasingly pertinent. In a well-known paper, Marx (2003) offers a typology of mechanisms of resistance to privacy invasions, ranging from completely passive, such as retreat from the surveillance-infused context, to active counter-moves such as masking one's identity. This paper starts from the premise that many of these strategies are no longer realistically attainable, and may indeed be counter-productive in the current digital environment.

This claim becomes evident when we consider the tension between privacy understood as an individual good, which decrees that consumers should be free to negotiate their acceptable levels of privacy, and privacy understood as indivisible collective value that can be enjoyed by the society only if a similar "minimum" level is afforded to every member (Regan, 1995). We argue that these two approaches to privacy should not be considered in opposition, but rather as supplementing each other. To advance the argument, we start by contrasting two possible strategies of individual resistance located at opposing poles of the trust placed in existing mechanisms of privacy protection—withdrawal from marketplace and trusting reliance on market-provided privacy protections—and discuss how such strategies may constrain privacy protection

understood as collective value. Then, we sketch an alternative approach to privacy that accounts for its collective nature and show how this approach may reorient current privacy protections.

Big Data and the Logic of Propensity

Despite the increasingly widespread references to big data, the term itself is ambiguous. Some refer to big data in a general sense, as an ecosystem composed of toolboxes (methodologies) and "mythologies" (boyd and Crawford, 2012) about the value provided to businesses and humanity by the large accumulation of data points. Others focus on the mechanics of collecting and storing large amounts of varied information. Gartner, Inc., a firm specializing in information technology (IT) research, provides one of the more common definitions: "Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making" (http://www.gartner.com/it-glossary/big-data). This definition encapsulates three aspects of big data environments: quantification, integration and processing. First, big data refers to the increasingly pervasive trend of "datafying" everyday life by transforming common activities into streams of data. The transformations of a great range of human activities (e.g., tweeting, running) and environmental signals into data produce large volumes of structured and unstructured datasets. A second aspect of big data refers to the infrastructure and expertise necessary for warehousing the ensuing information. Finally, big data refers to the development of algorithms to transform raw data into "actionable insights" that can allegedly help solve certain problems more efficiently by anticipating future states.

Thus, the emergence of big data as both a material and a discursive phenomenon marks what is identified as a profound societal change in the architecture, administration, and deployment of surveillance (Andrejevic, 2013). This change is most often imagined as a transition from panopticism—the central monitoring of pre-sorted bodies—to "panspectrocism" (De Landa, 1997)—a 360-degree, multi-sensorial monitoring of all human bodies administered by algorithms equipped with relevant watch-lists of "filters." This logic of surveillance, said to be characteristic for a post-disciplinary "society of control" (Deleuze, 1992), is presumably not so much about using visibility to reshape individuals (e.g., by forcing them to internalize punitive norms) as it is about deploying overlapping regimes of visibility to generate "shadows" of the individual. These "shadows" are eminently assessable replicas of the self, rendered "authentic" via the legitimacy offered by an array of new sciences (e.g., neuromarketing, data mining) and subject to observation, analysis and prediction outside the agency of the individual herself. As Deleuze (1992) puts it, the self as a formerly indivisible entity has become a "dividual" whose parts are modulated and recombined indefinitely according to criteria as unintelligible to the individual as they are uncontrollable.

It is in this sense that Thrift (2008) identifies the emergence of a principle of propensity as the organizing logic of the late capitalist marketplace. This logic works by deploying new technologies and disciplines to understand the inclinations of the customers so as to "nudge" them in the desired direction—a direction they are already susceptible toward—without the appearance of exercising control by constricting their choices. In what follows, we borrow Thrift's concept and take advantage of its dual meaning—an outside adjustment based on innate tendencies and a focus on statistically-

determined propensities—to capture two trends that increasingly characterize the operations of the commercial big data ecosystem and the nature of their power.

Propensity for Accumulation

First, "propensity" suggests a process of multiple adjustments to existing conditions by companies that aim to be on the cutting edge of marketing transformation. The key prerequisite of this unprecedented flexibility is the purposeful innovation of IT-based products that "mix and match" the streams of data accumulated through surveillance.

The market ideology of big data, rhetorically described as a "revolution" in how companies do business, invites a constant accumulation of data, the scaling up of the scope of use, and the repurposing of existing databases to provide additional intelligence services to a market primed to receive them. For example, Equifax, the second-largest consumer credit company, has extended beyond the financial area into identity verification, employment eligibility check, or assessing online social influence (Nash, 2012). This function creep requires that companies not only mine the "residue" of data that was collected for different purposes, but also start accumulating data beyond the original scope of the company, in order to be prepared for unanticipated data uses (Andrejevic, 2013).

To sell the logic of pervasive, cross-platform personal data collection to customers, companies increasingly emphasize the benefits offered by mass personalization, from heightened content relevancy that cuts through digital clutter to personalized recommendations that fit individual lifestyles. For example, Acxiom offers approximately half of its 1,500 data elements "to help marketers target consumers online

by personalizing websites for individual consumers or serving advertisements" (Committee on Commerce, Science & Transportation, 2013, 31). As this common marketing rhetoric would have it, persistent data collection is integral to a customer-centric approach that invites customers to "educate" their favorite brands on their preferences.

In many ways, consumers are complicit in this process. In order for the promised customer-centric personalization to work, users have to accept two premises. First, users need to accept the necessity of identifying themselves in ways that are not misleading about their preferences. For example, a user buying from Amazon for her entire family would not find the Amazon recommendation algorithm very useful, since the recommendations automatically generated would match the composite taste profile of all family members, rather than the preferences of that particular user. Second, users perceive recommendations systems to be more useful if they are personalized (Cremonesi, Garzotto and Turrin, 2012), which depends on whether users release enough information about their preferences.

The current impetus for data accumulation has invited external company restructuring in the form of mergers and acquisitions aimed at combining data silos. For example, to enhance its data mining capabilities, Equifax has spent $1.7 billion between 2007 and 2012 buying data-collection and technology businesses (Nash, 2012). Moreover, in terms of internal restructuring, companies are simultaneously creating new job positions that require specialized big data know-how in the form of novel data skills, and seeking integrative data analysis strategies that demand the unification of IT and marketing departments (Kennedy, 2012).

Propensity as Likelihood

In its second sense, "propensity" refers to statistical propensity, a logic of analysis that acts on the individual based on the probabilities of her category, rather than on her actual behavior (e.g., Andrejevic, 2013). Dubbed "decision analytics," these techniques of prediction of, and intervention into, individual behavior involve optimization of marketing via an increasingly fine-grained segmentation of consumers, context mining, real time identity verification, and feedback loops that allow the testing of marketing ideas on customers.

The appeal of data scientists supplanting creatives in marketing lies in their ability to provide "actionable insights" granted the authority of objective science without dealing with the vexing question of causality. The idea of data as value-generating rests on the assumptions that 1) big data inherently contains meaning, if only the right technologies and skill sets are mobilized to mine it; and 2) by increasing the scale of data collection to include as many consumer actions as possible, data becomes endowed with meaning simply waiting for the right algorithm to extract (Steel, 2012).

Big data analysis does not require scientists to engage in grounded research; indeed, their existence may be counterproductive. As Mayer-Schönberger and Cukier (2013: 64) claim, "[c]ontrary to conventional wisdom, such human intuiting of causality does not deepen our understanding of the world." Consequently, the use of data-rich predictive algorithms to anticipate consumers' behavior does not rely on extended knowledge of human nature or identification of behavioral causes, but on context triggers as indicated by correlation analysis (Andrejevic, 2013). As one journalist put it, "[t]he theory of big data is to have no theory, at least about human nature. You just gather huge

amounts of information, observe the patterns and estimate probabilities about how people will act in the future" (Brooks, 2013).

In many respects, this situation may evolve the "stoic neutrality of the social scientist and the dehumanizing character of bureaucracy" Andrejevic (2007: 68-69) by removing the "utmost" source of bias, the human one. Namely, the ideology of big data naturalizes algorithmic analysis of quantitative data as the paramount expression of truth (Cohen, 2013). Trade literature expresses this ideology by claiming that insights generated via big data need neither a priori knowledge, nor hermeneutic sensibilities: "People using big data are not like novelists, ministers, psychologists…coming up with intuitive narratives to explain the causal chains of why things are happening" (Brooks, 2013: A27). In other words, human agents are set to act only as data custodians and curators whose role is to correct pieces of information that may throw off a customer's profile. Hence, as Cohen (2013) argues, the ideological effect of big data is the denial of the existence of ideology and bias.

If, as cultural and social theorists have argued, the development of the self takes place at the intersection of social shaping and autonomous individuality, by prioritizing "algorithmic truth" over "human truth" the big data ideology not only implicates individuals' ability for self-making, but also constricts the flow of social influence and "environmental serendipity" (Cohen, 2013) that helps enrich identity.

It has been argued that the impetus to authenticate individuals in different contexts, which motivates the widespread use of various surveillance techniques (including biometrics), is inherently driven by a motivation to fixate and manage the multiplicity of identities into single ones (van Zoonen, 2013). In panoptic diagrams this

may particularly be so since, as Lyon (2007: 177) argues, multiple identities represent "a constant challenge to the would-be hegemonic system." Yet, in the panspectric diagram of big data, understanding individual behaviors requires that rather than seeing individuals, the surveillant entity focuses on the context (Palmås, 2011). Consequently, the threat that "algorithmic truth" poses on selfhood and individuals' ability to challenge the inferences made through big data analytics is not due to the imposition of a singular identity but rather the fleeting nature of the "truth" associated with the constantly recontextualized self. Below, for example, is how Equifax reconsiders notions of accuracy in big data environments:

> Where master data management projects seek the fabled single version of the truth, Brooks [senior executive at Equifax] says there's no such thing… "The reality is, they're all right. Now we think of observations more than truth" he says. (quoted in Nash, 2012, online)

Consequently, this understanding of human behavior creates pernicious power imbalances between institutions involved in big data analytics and individuals. First, while the circulation of information online can provide opportunities for individuals to correct "informational" inaccuracies, this ability nevertheless requires the distinction between accurate and inaccurate information (Solove, 2007). However, as suggested above, in the big data environment no such distinctions exist. Indeed, the only litmus test available for "truth" is whether the user's behavior ends up complying with the predicted pattern post-intervention (i.e., after being behaviorally targeted). As shown in a number of recent studies on virtual identity, this standard represents a very low barrier for "truth," since the prediction, and the consequent decisions regarding how to target the individuals,

are likely to create self-fulfilling prophecies. For example, research on what is called a "Proteus Effect" (e.g., Yee et al., 2009) suggest that even slight variations in the online identity (e.g., avatars in games) assigned to individuals may influence their behavior.

Second, big data analytics, which relies on stripping information from different contexts to create meaning, is antithetical to what Nissenbaum (2010) calls the "contextual integrity" of personal information. Accordingly, contextual integrity of information collapses when information meant to be used in one context, and hence governed by a particular set of norms and expectations about its flow, is used in other contexts.

Such collapses in the contextual integrity of information render individuals vulnerable to rational discrimination (Gandy, 2009). Services based on mining big data rely on metrics developed from cross-context databases that sort individuals into segments in terms of their desirability, identify risky segments that need to be excluded from services, and identify vulnerable segments that can be marketed to (Pasquale, 2015). Such sorting processes implicate the life chances of individuals by potentially altering individuals' future well-being (Gandy, 2009).

Privacy Self-Management: Market Avoidance vs. Market Assimilation

We discussed how the logic of propensity and algorithmic truth preempts data subjects' ability to challenge the inferences made about them by imposing ever-changing categories with fleeting claims to contextual validity. It is in this context that we need to reconsider the predominant neoliberal privacy regulatory framework that mandates individuals to self-manage their privacy behaviors via the "notice and choice" framework

(Solove, 2013) and places the onus on the individual to understand her risks and act accordingly.

In this framework, privacy carries individual value: it is the outcome of individuals' knowledgeable decision-making. Users are often defined as "privacy pragmatists" (Hoofnagle and Urban, 2014) who value their privacy yet can trade it in exchange for other benefits. At one pole, users may choose to reject the offered service if it does not come with the desired level of privacy. At the same time, to believe the marketing pundits, insofar as pragmatist consumers overcome their "irrational" fear that their data will be used against them and instead carefully examine the available options for protecting their data, they can participate fully in a mutually beneficial trade in personal information. Therefore, at the other pole, the "notice and choice" framework envisages a consumer who self-manages privacy by dealing strategically with data collection entities.

These assumptions are overwhelmingly illustrated in a diverse range of policy instruments regulating consumer data issued by both government bodies and industry alliances. For example, the "privacy by design" regulatory frameworks recently introduced in both U.S. and the European Union mandate individual responsibility for her privacy protection and understand individual empowerment as an enhanced ability to compare and contrast data practices (e.g., FTC 2014). Once properly equipped, consumers are expected to "evaluate their choices and take responsibility for the ones they make." (White House, 2012).

There is sufficient evidence to indicate that people are quite concerned about the use of their data (e.g., Turow et. al, 2015). More importantly, studies show that users'

actual knowledge about the mechanisms behind the new digital economy does not sufficiently equip them with the tools to protect their privacy. For example, users fail to read or understand privacy policies or to anticipate downstream data uses (Solove, 2013). In light of the discussion above which showed how the current big data environment handicap, from the start, users' ability to control the collection of data about themselves and challenge the meanings created from this data, these barriers to consumer education appear structural, rather than cases of inadequate individual capabilities.

Quite the contrary, the new literacies of the digital economy, particularly as envisioned by market-based entities, are not as much about increasing privacy risk awareness as they are about easing individuals into participating in the digital economy. For example, a representative from KBM Group, which proclaims to transform marketing into "mutually beneficial customer conversations through data-driven insights," argues that as part of customer education, the industry needs to explain "the value" created through the provision of better services in exchange for data (quoted in Dupré, 2013). Not surprisingly, "educating" the user in this context means creating a new market that capitalizes on consumers' privacy concerns by selling tools for the self-management of personal data.

Take Abine's costly privacy-enhancing product DeleteMe which claims to "delete your personal information from the Internet," by which the company means deleting "your public profile from leading data sites; contact, personal, and social information; and photos of you, your family, and your home." The unique selling proposition of DeleteMe is very clear: it will protect individual privacy by removing data from (a limited number of) public databases and keep personal data away from the leering eyes of

"evil-doers". For the average consumer, the premise is hard to resist. Yet, the pervasiveness of data collection and warehousing, the use of automated analytics for producing inferences about people, and the non-transparent way in which this infrastructure operates cast doubt on the claims of services like DeleteMe. Their limitations also echo a larger problem with the assumption that equipping individuals with the knowledge and the tools to protect their privacy will be sufficient to provide users with meaningful options.

Moreover, as illustrated below, the "notice and choice" framework, with its focus on individual as the locus of privacy decisions, may not only fail to protect individual privacy, but also bias the privacy calculus of the larger society by reducing the level of privacy available to all. In what follows, we discuss the long-term outcomes associated with either withdrawing from the market as a form of resistance through avoidance, or completely relying on market products for modulating individual privacy levels. Our discussion will help illustrate how these two different strategies that cater to privacy needs at micro (i.e., individual choice) levels impact collective privacy values by reducing the level of privacy attainable. Later in the paper, we will propose an alternative conception of privacy that focuses on its collective dimension.

Market Avoidance: The Awareness Paradox

In a recent article on the development of a new information literacy curriculum, Cirella (2012) proposes an "account-based literacy" program to help increase users' awareness of how "the end result [of sharing bits of information] may be a much fuller view of users' personal lives than they consciously disclosed in each separate setting." Accordingly, only through such literacy can users actively determine what information to

share. Yet, to the extent that the market is dominated by privacy policies that adopt a "take-it-or-leave-it" approach wherein users can either consent to data collection or are deprived of critical services (Popescu and Baruh, 2013), we need to question whether users' knowledge of surveillance and data mining practices can translate into meaningful actions.

One possible user response to the unavoidable creation and possible misuse of users' digital footprints may result in what can be named an *awareness paradox*. Namely, as Kosinski et al. (2013) predict, heightened awareness of how inferences are made about individuals, along with the difficulty of controlling which behaviors are revealed and what inferences can be made from them, can "completely deter [consumers] from using digital technology." In other words, awareness may potentially lead to complete or partial withdrawal from the digital marketplace as a tactic of resistance.

Recent evidence indicates that although still small in size, there may be a growing group of users who, as a result of their heightened awareness of the impossibility of protecting their privacy, will either withdraw (however partially) from the "digital market" or not enter it in the first place. It is virtually impossible to completely withdraw from the online marketplace for privacy reasons because having an online presence (for example, owning a credit card) is a necessary part of accessing a wide variety of consumer services (Wessels, 2015). Nevertheless, at least in respect to the use of social media, partial withdrawal for privacy reasons, for example by limiting engagement with social networks like Facebook (Staddon, Huffaker and Brown, 2012) or even deleting one's Facebook account (Baumer et. al, 2013), are increasingly visible practices (Foot, 2014). Dubbed "virtual identity suicide" and promoted by campaigns (e.g., the 2010

"Quit Facebook Day") and art projects (e.g., Seppukoo.com and Web 2.0 Suicidemachine), these initiatives invite users to commit "digital suicide" from Facebook both as a tactic of disruption (Karppi, 2011) and as a way of life (Portwood-Stacer, 2012). A recent empirical study of Facebook quitters confirmed that privacy concerns are among the top factors influencing the decision to quit: 48% in a sample of 310 Facebook quitters cited privacy concerns as the main reason for closing their accounts (Stieger et al., 2013). These practices go beyond Facebook. For example, a survey conducted by the Oxford Internet Institute in the U.K. shows that for both internet "ex-users" (about 5% of the population) and "non-users" (about 18% of the population), privacy concerns were among the primary reasons for not going online (Dutton and Blank, 2013: 6). Although in its infancy, the field of study of non-users confirms the existence of an important segment of non-users of various digital services who are highly intolerant of privacy violations (e.g., Bright, Kleiser, and Grau, 2015).

This withdrawal has several paradoxical consequences. First, if we were to adhere to the dominant conceptualization of privacy as a "commodity" with a market-determined value, the withdrawal of the privacy-risk-aware segment would reduce the availability of signals that companies are supposedly using to determine privacy preferences. Second, because the consumers who value their privacy too much to trade it for benefits are no longer in the market, the marketplace creates a self-fulfilling prophecy that all consumers are privacy pragmatics who use their privacy as a token of exchange. This skew in existing privacy preference signals reduces market incentives to cater to the privacy needs of those segments.

Finally, even for the users who choose to stay in the market, awareness of surveillance practices may steer their usage patterns in ways that do not reflect the extent to which those users would protect their privacy if given a meaningful choice. The concept of "privacy paradox" has been frequently used to discuss the discrepancy between users' declared concern about their privacy and their allegedly careless online behaviors (e.g., Barnes, 2006). However, recent empirical evidence (e.g., Turow et al., 2015) suggests that one of the main reasons for the privacy paradox is the lack of a meaningful privacy choice resulting in an attitude of "Why bother?".

Although these examples may be different in respect to outcomes, they affect market signals similarly. That is, the data that the market relies on to determine users' privacy needs underestimate the level of privacy expected by the users, which, in the long run, may reduce the general level of privacy for all.

Market Assimilation: Privacy Aware Systems

The concept of awareness paradox underlines how accumulation of breaches in individuals' trusts may result in a self-fulfilling prophecy wherein users' ability (and willingness) to signal their privacy preferences to the market diminishes. However, it is also possible to envision a future within which markets will be able to use "privacy aware systems" that adjust to individual privacy preferences (Sackmann, Strüker, and Accorsi, 2006) by negotiating the point of equilibrium between the amount of data disclosure individuals may accept and the level of benefits in digital services.

At first blush, these systems address the problem of restoring user bargaining power in the privacy marketplace by allowing the automatic customization of user privacy preferences (Popescu and Baruh, 2013). However, as the remainder of this

section will show, this form of privacy negotiation may actually result in new forms of "privacy inequities" whereby companies take advantage of users' lack of knowledge about privacy protections to negotiate privacy rights differentially, thus creating a class of "privacy have-nots" who perform "data labor" disparately.

The idea of privacy-aware systems is not new. For example, protocols such as P3P or Xpref (developed in early 2000s) aimed to allow both organizations and users to set up their data collection and privacy preferences (e.g., the kind of data that can be collected and who can have access to it) and then match user preferences with the organizations' privacy policies (De and Pal, 2013; Sackmann et al., 2006). However, such protocols have been marred by a number of problems, such as the difficulty of implementation, given the number of entities with whom users interact or the time and effort required for users to set them up, that prevented their widespread adoption (Corapi et al., 2009). Additionally, these protocols rely on default privacy policies that often provide users with "all or nothing" options (e.g., accept policies or not receive a service) that are not responsive to context-dependent privacy needs (Bigwood et al., 2012).

A newer framework known as adaptive privacy settings is said to solve problems faced by privacy specification schemes that rely on default privacy policies (Bigwood et al., 2012). Adaptive privacy settings are systems that utilize machine-learning techniques to extract individuals' privacy preferences and configure their privacy settings. Unlike previous privacy specification schemes, setting up adaptive privacy systems would require less time and effort from the users and provide an intelligent customization of privacy preferences based on temporal and locational aspects of and the parties involved in an interaction (Bigwood et al., 2012; Corapi et al., 2009).

In addition to being customizable based on user demographics, adaptive privacy systems are also predictive. For example, for mobile communications, Corapi et al. (2009) use a combination of locational, temporal, and behavioral (i.e., accept/reject incoming calls) data, along with classification of the source of the call (e.g., friend, home, college) to predict user behavior for incoming calls. Likewise, for social network sites, users' profile elements, list of friends, size of friends' list, privacy preferences of friends in a users' community, object meta-data (e.g., photo tags on Facebook) have been used to predict privacy preferences of users (Banks and Wu, 2010; Bigwood et al., 2012). Recently, different proposals have sought to enhance adaptive systems by using semantic analysis of textual data (Li et al., 2011). These techniques have met with some success. Studies report that current machine-classification systems can reach more than 80% accuracy in predicting privacy preferences of individuals (Li et al., 2011; Toch, 2014) and decrease over-exposure of personal information by 40% (Bigwood et al., 2012).

Yet, aside from the irony that increasing the efficiency (accuracy per user effort) of adaptive privacy systems requires the collection of even more finely-grained behavioral data, adaptive privacy systems may have important implications for the relationship between user choice and privacy protection in big data environments. First, the underlying rationale for adaptive privacy systems is very similar to that of automated product recommendation systems, namely, conserving individual cognitive and time resources. The fact that users have bounded rationality means that they lack both the capacity and the will to engage in a complicated privacy calculus with all data-handling agents (Banks and Wu, 2010). However, it should be noted that the existence of bounded rationality also actualizes the potential of automated recommendation systems (or in our

case adaptive privacy systems) to not only detect preferences, but also construct them, particularly when users do not have well-formed and stable privacy preferences (Adomavicius et al., 2011).

Second, while the intended use of the adaptive privacy systems described in this section is to assist individuals in protecting their privacy, adaptive systems can also be a vital component of a marketplace in which privacy becomes increasingly commodified as a monetized service rather than a value. For example, Tian et al. (2011) argues that as digital services increasingly move to cloud computing, consumers can enjoy different levels of privacy based on their ability to pay and the cost of supplying that service. Within this kind of a marketplace for privacy protection, adaptive privacy systems, which currently aim to reduce "over-exposure," can also work by minimizing "under-exposure."

While widespread adoption of adaptive privacy systems using big data analytics for differential privacy treatments may be seen as a distant possibility, the impetus to do so is already present. For example, Zhou and Piramuthu (2014) propose that companies use algorithms similar to those employed in personalized pricing to customize the level of data protection and transparency afforded to users. The Simmons Privacy Segmentation system that Experian offers to enterprises divides users into six different clusters as a function of their sensitivity to data privacy invasions (Experian Marketing Services, 2013). More specifically, current patent applications as well as trade/scientific literature point to the pending development of privacy auction systems wherein each "violation of privacy" will be priced differentially depending on how much individual users value their privacy (Ghosh and Roth, 2011; Ioannidis et al., 2013).

This scenario would represent a form of privacy inequity wherein machine-

learning techniques borrowed from big data analytics are used to determine how much personal information can be collected from the data subject without leading to user churn (or, alternatively, how much privacy is worth to each individual). In some respects, this privacy differential corresponds to what Ericson and Haggerty (2006: 12) describe as a new politics of surveillance entailing "inducements and enticements at the precise threshold where individuals will willingly surrender their information." In other respects, this situation introduces privacy inequities insofar as it further cements the nature of privacy as a tradable good: since companies recoup the costs of "missing out" on the personal information of privacy-conscious individuals by increasing the price of service, the individual level of privacy protection will increasingly depend on users' ability to pay.

Privacy as a Collective Phenomenon

The section above contrasted two tactics of user resistance to ubiquitous data mining that fit with the logic of individual privacy self-management, and argued that both would result in the underestimation of the value of privacy, if not outright privacy inequity. This seemingly cynical conclusion might suggest that, in Borg-like fashion, "resistance is futile." In this section we argue, however, that an alternative understanding of privacy can help create new avenues of privacy protection.

Privacy approaches could be classified on a individualism-collectivism continuum that identifies where the locus of responsibility lies for social action. As per Aaker and Williams (1998), collectivist cultures promote the collective ("us") over the individual ("me"), are characterized by the recognition of human interdependency, and are concerned with "blending the self and the other boundary" (p. 242). In this vein, to think

of privacy from the perspective of the collective prompts the recognition of privacy both as collective value (codependency) and as collective social phenomenon (cooperation).

Privacy codependency pertains to privacy as a "collective value" (Regan 2015), meaning that the available level of privacy in a given context depends not only on individual isolated choices, but also on the choices of other societal agents (other individuals or institutions). As collective value, privacy cannot be enjoyed by a person without "all persons having a similar minimum level of privacy" (Regan, 1995: 213). Moreover, within existing socio-technical platforms, individuals may not achieve custom levels of privacy outside the constraints created by the platform's technical specifications and its users' behaviors (Regan, 2015). Illustrative of privacy codependency, recent evidence suggests that even for non-users of a social network, the platform may still employ big data techniques to infer "full shadow profiles" about them based on data collected from their contacts that are using the network (Sarigol, Garcia, and Schweitzer, 2014).

Importantly, the implication of privacy codependency is that privacy harms happen not only at individual, but also at collective level, just like one incident of oil spill may affect the entire environment (Hirsch, 2014). For example, as Hirsch argues, the frequency of data breaches that involve considerable amounts and variety of data (e.g., the recent data breach of 80 million records at Anthem Inc., the second largest health insurer in the U.S.) alter the perceived risk of identity theft, which results in increased costs for both companies ("clean up" costs) and consumers (the cost of investing in prevention tools).

Beyond its value for the collective, recent research is starting to acknowledge that

privacy is not an isolated phenomenon, but rather a collective social phenomenon that involves cooperation. In that respect, privacy management entails what Petronio (2002) describes as a sharing of the responsibility of protecting boundaries. While discussions of "networked privacy" (Marwick and boyd, 2014) allude to how privacy may emerge from the negotiation of mutually adjustable boundaries, cooperation does not necessitate a consensus on a single set of privacy norms. Rather, as Martin (2012) argues, cooperation implies the recognition that while individuals' privacy priorities may vary, their privacy expectations need not diminish.

Indeed, recent evidence indicates that cooperative thinking is part of individual decision-making about privacy. For example, in a cluster analysis of online users' privacy attitudes, Baruh and Cemalcılar (2014) identified a large segment of "privacy advocates" who voiced concern not only for their own privacy, but also for the "privacy of others." Notably, this segment was more likely than "privacy individualists" to engage in privacy-protective measures such as minimizing their own disclosure. These findings illustrate not only that for a considerable proportion of users, decisions regarding whether or not to share information are often influenced by both concern about one's own privacy and concern about privacy of others (Baruh and Cemalcılar, 2014), but also that in networked environments users may often employ group privacy management approaches (De Wolf et al., 2014).

Acknowledging the collective aspect of privacy permits alternative approaches to both privacy risk assessment and the ensuing remedies. Despite emphasizing the potential harms to the society at large, privacy assessment tools (PIA) typically prompt organizations to focus on three categories of risks: risks to the immediate data subjects

(i.e., the subjects of data collection and processing), risks to the organization (e.g., losing trust, incurring "clean up" expenses), and risks from non-compliance with regulations (Information Commissioner's Office, 2014). Alternatively, under the privacy as collective phenomenon perspective, PIA would also aim to account for what Sarigol et al. (2014) define as the "privacy leakage factor," namely the extent to which the actions of specific data subjects produce privacy loss to others. Moreover, accounting for the cooperative nature of privacy means informing individual decision-makers not only about their individual privacy risks, but also about the privacy risks their actions pose to others. Existing literature on health interventions targeting behavior that pose risk to self and others (e.g., smoking or intravenous drug use) suggests that this approach may not only reduce the gap between attitudes and behavior (here, the privacy paradox), but also have more sustainable effects because they appeal to individuals' ego-related commitments (e.g., altruism) (Bethea, Murtagh and Wallace, 2015; Friedman et al. 2015).

Conclusion

In the Western tradition, privacy is often seen as a point of individual resistance against the increasingly pervasive surveillance by the state and commercial actors. By encouraging individual autonomy (Gavison, 1980), privacy supposedly protects individual power of self-determination and, not least, individual capacity for self-definition. In a pervasive data collection environment, privacy discussions are complicated by the alleged difficulty of defining harm from privacy violations at the individual level. Quite the contrary, companies claim, users now expect real-time hyper-personalization of their online environments, and relinquishing privacy in exchange is mutually understood as the default for such services. According to this logic, instead of

being a right, privacy becomes the price consumers have to pay for being addressed as autonomous actors rather than predictive categories.

Yet, as we showed in this paper, the algorithmic social sorting characteristic of big data environments drastically limits the ability of individuals to self-define, and thus claim control and agency, over their social trajectory. Surveillance processes that work to thin-slice populations into abstract, algorithmically-produced categories are not only far removed from the "selfhood categories" individuals might use to define themselves, but also recontextualize the self in a fleeting and unchallengeable manner. Indeed, the "context is truth" mantra adopted by many companies translates into the constant redefinition of commercial categories of intervention, based on the accumulation of new streams of data and real-time contextual triggers. The process of classifying individuals becomes a constantly updated commercial "black box" (Pasquale, 2015) that they are unable to challenge. Thus, the ideological power of the big data logic is to render the forces that shape decisions over individual lives both ubiquitous and unintelligible to the individual.

Regulatory efforts in the past several years have sought to update privacy protections to address data collection in the digital environment, all the while retaining the "notice and choice" framework that assumes a knowledgeable and privacy-conscious user (Solove, 2013) who pursues her privacy in isolation. We argued that such assumptions effectively rationalize market withdrawal for the privacy-conscious individual (the Awareness Paradox), while creating new power imbalances for the individuals that fully rely on the market-produced solutions. The withdrawal, however partial, from the market of those individuals highly intolerant of privacy violations only

serves to further skew market signals by legitimizing the argument that "digital natives" have different, laxer privacy expectations. Reliance on the market-situated, (neoliberal) individual privacy management transforms privacy from a consumer right into a consumer service, to be modulated and personalized based on data collection about a user's specific privacy preferences. Indeed, as shown above, the existing technologies to automate privacy preferences create differential privacy protections and consequently a digital literacy-driven privacy "discrimination" whereby the personal data of the privacy "have nots" is disparately and more extensively exploited.

As long as regulatory efforts center on individual privacy literacy and self-management but fail to recognize the nature of privacy as both a collective value and a collective-social phenomenon, these efforts are destined to fail while leaving precious few avenues of resistance to individuals. Instead, we suggest that restoring genuine user agency means designing protections and remedies that specifically acknowledge the collective aspect as privacy. Such remedies may involve new ways of calculating privacy risks, as well as new ways to frame risk information to expand the benchmark used by individual decision-makers. Also, a potential value of a recognizing the collective-social dimension of privacy lies in its ability not only to challenge the assumptions of dominant market-based models of privacy protection, within which "individual" privacy concerns are at the losing end of a balancing act that pit them against larger goals as market efficiency but also to create avenues through which individual acts of resistance may transform into "mass sensibility for spontaneous resistance" (Bennett, 2008: 212) on a collective scale.

References

Aaker JL and Williams P (1998) Empathy versus pride: The influence of emotional appeals across cultures. *The Journal of Consumer Research* 25(3):241–261

Abine (2015). Delete your personal information form the Internet. Available at: https://www.abine.com/deleteme/landing.php.

Adomavicius G, Bockstedt J, Curley S, et al. (2011) Recommender systems, consumer preferences, and anchoring effects. In *Proceedings of the 5th ACM Conference on Recommender Systems,* pp.35-42. New York: ACM.

Andrejevic M (2007) *iSpy: Surveillance and Power in the Interactive Era.* Lawrence: University Press of Kansas.

Andrejevic M (2013) *Infoglut: How Too Much Information is Changing the Way We Think and Know.* New York: Routledge.

Banks L and Wu S (2010) Toward a behavioral approach to privacy for online social networks. In: Bolc L, Makowski M and Wierzbicki A (eds) *Social Informatics.* Berlin: Springer Berlin Heidelberg, pp.19-34.

Barnes SB (2006) A privacy paradox: Social networking in the United States. *First Monday* 11:11-15.

Baruh L and Cemalcılar Z (2014) It is more than personal: Development and validation of a multidimensional privacy orientation scale. *Personality and Individual Differences* 70(November): 165-170.

Baumer EPS, Adams P, Khovanskaya VD, et al. (2013) Limiting, leaving, and (re) lapsing: An exploration of Facebook non-use practices and experiences. In: *Proceedings of the SIGCHI Conference*, pp.3257-3266. New York: ACM.

Bennett, CJ (2008) *The Privacy Advocates: Resisting the Spread of Surveillance*. Cambridge: The MIT Press.

Bethea J, Murtagh B and Wallace SE (2015) "I don't mind damaging my own body" A qualitative study of the factors that motivate smokers to quit. *BMC Public Health* 15(4):1-9.

Bigwood G, Abdesslem FB and Henderson T (2012) Predicting location-sharing privacy preferences in social network applications. In: *Workshop on Recent Advances in Behavior Prediction and Pro-active Pervasive Computing.* Newcastle, UK, 19 June 2012.

boyd d and Crawford K (2012) Critical questions for Big data. *Information, Communication and Society* 15(5):662-679.

Bright LF, Kleiser SB and Grau SL (2015) Too much Facebook? An exploratory examination of social media fatigue. *Computers in Human Behavior* 44:148-155.

Brooks D (2013) What you'll do next. *The New York Times,* 16 April, p.27.

Cirella D (2012) Beyond traditional literacy instruction: toward an account-based literacy training curriculum in libraries. *Computers in Libraries* 32(10):5-9.

Cohen JE (2013) What privacy is for. *Harvard Law Review* 126(7):1904-1933.

Corapi D, Ray O and Russo A (2009) Learning rules from user behaviour. In: Iliadis L, Vlahavas I and Bramer M (eds) *Artificial Intelligence Applications and Innovations III.* Berlin: Springer Berlin Heidelberg, pp.459-468.

Cremonesi P, Garzotto F, and Turrin, R. (2012). Investigating the persuasion potential of recommender systems from a quality perspective. *ACM Transactions on Interactive Intelligence Systems* 2(2): 1-41.

De Landa M (2000) Deleuze, diagrams and the genesis of form. *American Studies* 45(1): 33-41.

De SJ and Pal AK (2013) Cloud-based privacy aware preference aggregation. In: Cuzzocrea A, Kittl C, Simos DE, et al. (eds) *Availability, Reliability, and Security in Information Systems and HCI.* Berlin: Springer Berlin Heidelberg, pp.208-223.

De Wolf R, Willaert K and Pierson J (2014) Managing privacy boundaries together: Exploring individual and group privacy management strategies in Facebook. *Computers in Human Behavior* 35:444–454.

Deleuze G (1992) Postscript on the societies of control. *October* 59:3-7.

Dixon P, and Gellman R (2014) *The scoring of America: How secret consumer scores threaten your privacy and your future*. World Privacy Forum. Available at http://www.worldprivacyforum.org/2014/04/wpf-report-the-scoring-of-america-how-secret-consumer-scores-threaten-your-privacy-and-your-future/.

Dupré E (2013) Top 5 privacy issues revealed. *Direct Marketing News,* 1 February. Available at: http://www.dmnews.com/top-5-privacy-issues-revealed/article/277683/#

Dutton WH and Blank G (2013) Cultures of the internet: The internet in Britain. Oxford Internet Survey, UK.

Ericson, RV and Haggerty KD (2006). The new politics of surveillance and visibility. In Ericson RV, Haggerty KD (eds) *The New Politics of Surveillance and Visibility*. Toronto: University of Toronto Press, pp.3-25.

Experian Marketing Services (2013) Consumer Insights. Available at: http://www.experian.com/simmons-research/online-privacy-segmentation.html.

Federal Trade Commission (2014) *Data Brokers: A Call for Transparency and Accountability*. Available at: https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

Foot, K (2014) The online emergence of pushback on social media in the united states: a historical discourse analysis. *International Journal of Communication* 8:1313-1342.

Friedman, S.R. et al., 2015. Measuring altruistic & solidaristic orientations towards others among people who inject drugs. Journal of Addictive Diseases, (June 2015),

Gandy OH (2009) *Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage,* Burlington, VT: Ashgate.

Gavison, R (1980). Privacy and the limits of law. *Yale Law Journal*, 89(3):421-471.

Ghosh A and Roth A (2011) Selling privacy at auction. In: *Proceedings of the 12th ACM Conference on Electronic Commerce*, pp.199-207. New York: ACM.

Hirsch DD (2014) Glass house effect: big data, the new oil, and the power of analogy. *Maine Law Review* 66(2):373-395.

Hoofnagle CJ and Urban JM (2014) Alan Westin's privacy homo economicus. *Wake Forest Law Review* 49:261-309.

Information Commissioner's Office (2014) *Conducting Privacy Impact Assessments Code of Practice*. Available at https://ico.org.uk/media/for-organisations/documents/1595/pia-code-of-practice.pdf

Ioannidis S, Fawaz N and Dandekar P (2013) *Privacy Auction Mechanism*, US Patent PCT/US2012/059302

Karppi T (2011) Digital suicide and the biopolitics of leaving Facebook. *Transformations* (20) (accessed June 28 2015) http://www.transformationsjournal.org/journal/issue_20/article_02.shtml.

Kennedy J (2012) The big data dilemma. *B&T Weekly*, 3 September. Available at: http://www.bandt.com.au/features/the-big-data-dilemma.

Kosinski M, Stillwell D and Graepel T (2013) Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America* 110(1):2-5.

Li Q et al. (2011) Semantics-enhanced privacy recommendation for social networking sites. In: *Proceedings of the 10th International Conference on Trust, Security and Privacy in Computing and Communications*, pp.226-233. Los Alamitos: IEEE.

Lyon D (2007) *Surveillance Studies.* Cambridge: Polity Press.

Marwick AE and boyd d (2014) Networked privacy: How teenagers negotiate context in social media. *New Media and Society* 16(7):1051-1067.

Marx GT (2003) A tack in the shoe: Neutralizing and resisting the new surveillance. *Journal of Social Issues* 59(2):369-390.

Mayer-Schönberger V and Cukier K (2013) *Big data: A Revolution that will Transform How We Live, Work, and Think.* Boston: Houghton Mifflin Harcourt.

Nash KS (2012) Equifax eyes are watching you—Big data means Big Brother. *CIO,* 15 May. Available at: http://www.cio.com/article/706457/Equifax_Eyes_Are_Watching_You_Big_Data_Means_Big_Brother.

Nissenbaum H (2010) *Privacy in Context: Technology, Policy, and the Integrity of Social*

Life. Stanford: Stanford University Press.

Palmås K (2010) Predicting what you'll do tomorrow: Panspectric surveillance and the contemporary corporation. *Surveillance & Society* 8:338-354.

Pasquale F (2015). *The Black Box Society*. Cambridge: Harvard University Press.

Petronio SS (2002). *Boundaries of Privacy: Dialectics of Disclosure.* Albany: State University of New York Press.

Popescu M and Baruh L (2013) Captive but mobile: Privacy concerns and remedies for the mobile environment. *The Information Society* 29 (5):272–286.

Portwood-Stacer L (2012) Media refusal and conspicuous non-consumption: The performative and political dimensions of Facebook abstention. *New Media and Society* 15(7):1041-1057.

Regan P (1995) *Legislating privacy: Technology, social values and public policy.* Chapel Hill: University of North Carolina Press.

Regan P (2015), Privacy and the common good: revisited. In Roessler B, Mokrosinska D (eds), *Social Dimensions of Privacy: Interdisciplinary Perspectives.* Cambridge; Cambridge University Press, pp.50-70.

Sackmann S, Strüker J and Accorsi R (2006) Personalization in privacy-aware highly dynamic systems. *Communications of the ACM* 49:32-38.

Sarigol, E., Garcia, D., & Schweitzer, F. (2014). Online privacy as a collective phenomenon. In *Proceedings of the Second Edition of The ACM Conference On Online Social Networks,* pp.95-106. New York: ACM.

Senate Committee on Commerce, Science, and Transportation. (2013). *A review of the data broker industry*.

http://www.commerce.senate.gov/public/?a=Files.Serve&File_id=0d2b3642-6221-4888-a631-08f2f255b577

Singer, B. N. (2012). You for sale. *The New York Times,* 17 June, p.BU1.

Solove DJ (2007) *The Future of Reputation: Gossip, Rumor, and Privacy on the Internet*. Caravan Books.

Solove DJ (2013) Privacy self-management and the consent dilemma. *Harvard Law Review* 126(7):1880-1902.

Staddon J, Huffaker D, Brown L et al. (2012). Are privacy concerns a turn-off? Engagement and privacy in social networks. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, pp.1-13.

Steel E (2012) Data scientists take byte out of Mad Men. *Financial Times*, 12 December. Available at: http://www.ft.com/intl/cms/s/2/db8d250e-4279-11e2-979e-00144feabdc0.html.

Stieger S et al. (2013) Who commits virtual identity suicide? Differences in privacy concerns, Internet addiction, and personality between Facebook users and quitters, *Cyberpsychology, Behavior and Social Networking* 16(9):629-34

Thrift N (2008) Pass it on: Towards a political economy of propensity. *Emotion, Space and Society* 1(2):83-96.

Tian Y, Song B and Huh E (2011) A flexible cost-based privacy-aware data integration system in cloud. In: Park JJ, Yang LT, Lee C (eds) *Future Information Technology.* Berlin: Springer Berlin Heidelberg, pp.387-396.

Toch E (2014) Crowdsourcing privacy preferences in context-aware applications. *Personal and Ubiquitous Computing* 18(1):129-141.

Turow J, Hennessy M, and Draper N. (2015). *The tradeoff fallacy: How marketers are misrepresenting American consumers and opening them up to exploitation*. Available at https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf

van Zoonen L (2013) From identity to identification: fixating the fragmented self. *Media, Culture & Society* 35(1):44-51.

Wessels B (2015) Authentication, status, and power in a digitally organized society. *International Journal of Communication* 9:2801–2818.

White House (2012). Consumer data privacy in a networked world: A framework for protecting privacy and promoting innovation in the global digital economy. Available at: http://www.whitehouse.gov/sites/default/files/privacy-final.pdf

Wikibon (2014). Big Data Vendor Revenue and Market Forecast 2013-2017 http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2013-2017

Yee N, Bailenson JN and Ducheneaut N (2009) The Proteus effect: Implications of transformed digital self-representation on online and offline behavior. *Communication Research* 36(2):285-312.

Zhou W and Piramuthu S (2014) Information relevance model of customized privacy for IoT. *Journal of Business Ethics* (accessed 15 June 2015) http://link.springer.com/article/10.1007%2Fs10551-014-2248-y#