Privacy Across Economic Contexts

Sebastian Benthall, Cornell Tech

November 17, 2017

Abstract

It has been argued that the economic study of privacy must be fragmented across the many economic contexts which use personal information. This paper argues that contextually fragmented economics of personal information will not capture the most pressing privacy problem, which is the unpredictable secondary use of personal information in contexts that it was not originally collected in. The statistical properties of information flow in causal systems imply that the range of potential secondary uses will be unknown to consumers. This information assymetry due to the empirical complexity of understanding the consequencs of personal information disclosure suggests the conditions of market failure. Two-sided markets are identified as a way for firms to capitalize on the multiple purposes of personal data by renting out associational information through advertising services. As a solution to the implied market failure, we consider data protection regulation similar to that already in place in other expert service sectors (e.g. health, finance, and education).

1 Privacy and economics

It is well known that privacy is a many-meaninged term difficult [26] if not impossible [12] to pin down. Modern economics of privacy concerns itself mainly with concerns about the economics of personal information as it is represented and used by businesses employing information technology. Specially, it most often addresses what Acquisti, Taylor, and Wagman [1] call *tangible* impacts of privacy, those impacts that have objectively measurable and modelable costs and benefits and effects on market structure. While others acknowledge the possible importance of *intangible* impacts, such as a psychological concern about how ones personal information may be used (which may be modeled as a subjective preference for privacy [4] [5]) and

other more global social effects, we will limit the discussion in this paper to tangible impact.

Even so narrowly scoped, there are many different economic contexts in which the presence or absence of personal information is critically relevant. There are so many different contexts, each represented in their own sophisticated scholarly literatures, some [1] argue that a comprehensive economics of privacy cannot be achieved. This particularist approach cannot, however, address what may be a critical challenge of privacy economics, which is to address the economic impact of information flowing *between economic contexts*.

To be more concrete, it is well known that personal information is used by on-line retailers for price differentiation [24, 27]. With differentiated prices, a firm can charge consumers closer to their individual reservation price instead of choosing one price for all consumers. This has the effect of reducing deadweight loss by charging consumers with very low willingness to pay a price they can afford, while transforming consumer surplus formerly accrued by those with high reservation price to the firm as producer surplus. This is an example of tangible impact of privacy in economics.

While classically price differentiation can be based on personal information that is, by superficial judgment, directly relevant to a consumer's reservatin price, such as prior purchasing history [2], a problem for the economics of privacy is that theoretically *any* personal information may have some bearing on ones reservation price. This opens up price differentiation as a *secondary use* of any personal information that is collected or volunteered for any other purpose.

A similar argument may be made for markets where natural persons are in some sense the product being sold, as is the case of the labor, insurance, and credit markets. In each of these markets a firm must evaluate natural persons for their individual capacities (to perform a certain kind of work, to avoid risk, or to repay a loan) and decide whether to invest resources in them. The firm generally benefits from having more information about the persons under consideration. The effect of privacy, or lack of it, is uneven across the population being considered by the firm. Paradigmatically, more suitably employees are benefited if their suitability is known to potential employers, while conversely less suitable employees are harmed by the same. [19]

While it has been argued that more efficient sorting of individuals in these markets leads to economic surplus, this may be little consolation to those negatively impacted by the flow of their personal information. In this narrowly considered economic context there is a global tradeoff between economic productivity, lubricated by flows of personal information, and equal-

ity. Moreover an individual, and especially more vulnerable individuals, may be negatively impacted in these markets for human capacity by secondary uses of personal information collected in other contexts.

It is possibly because of the potential negative impact of secondary uses of information that so many market segments are protected by sectoral privacy laws. HIPAA in health care, GLBA in personal finance, and FERPA in education all place restrictions on firm's ability to disclose personal information. What these sectors all have in common is that they do not function without significant disclosures of personal information to firms because the provide personalized service unavailable to the consumer. We will use the term *expert services* to refer to this class of markets.

The more an expert service depends on the disclosure of intimate or sensitive detail, the more highly restrictive the confidentiality requirements associated with the service are. For example, HIPAA has special provisions for psychotherapy notes that do not apply to personal health information more generally. Attorney-client privilege, which protects personal information disclosed to ones lawyer, is another example of strongly protected confidentiality [{**?**]hazard1978historical, allen1990positive, richards2007privacy}. Noting that without confidentiality, personal information disclosed to expert services could have harmful tangible impact through secondary use in other markets is a reason to include it in consideration of privacy economics. It is notable that in all these cases of expert services, data protection is mandated by law, not left for market selection or self-regulation.

The economics of privacy and the economics of security intersect at the use and misuse of personal information. One the one hand, we can consider security to be another context where personal information is used, perhaps in a secondary way. Uses of personal information which are harmful to all affected consumers include those that facilitate security threats like spearphishing (when attackers use personal information to manipulate a person to reveal security-related information or otherwise be a vector for a furhter attack) and identity theft. On the other hand, it is the possibility of harmful secondary use *across all potential contexts* that makes security of personal information so important in the first place. Security in this sense is necessary for an implementation of confidentiality.

2 Privacy, context, and information flow

We take as a known result that consumer privacy expectations and preferences are tied to social context. Here, we follow Contextual Integrity (CI)

[14], a theory of privacy which maintains that public outcry over violations of privacy is due to violations of socially entrenched norms about the appropriate flow of personal information. According to the theory, these norms inhere in differentiated social contexts or spheres, such as the family, education, health care, finance, and so on. This theory has been confirmed empirically [11] through tests of consumer preferences: indeed consumer expectations of privacy vary with social context, perhaps explaining why coarser grained studies of privacy preferences are inconsistent with user's actions. In addition to being well grounded in legal literature and social scientific theory and validated by empirical research, CI has been influential in privacy engineering [25].

There are two upshots of CI that we highlight here. The first is that if consumer expectations of privacy are sensitive to social spheres with particular roles and information types, then they will be ill-equipped to comprehend the implications of secondary information use in contexts other than the one where the data was collected. Indeed, expectations of confidentiality may be essential for reducing the cognitive load of understanding the effects of information disclosure because to consider all the possibilities is literally impossible for an individual. Evidence that consumers are generally not making privacy decisions in rational and informed self-interest but, rather, become much more concerned with their privacy when told facts about how personal information is used [10] shows that there is a disconnect between consumer expectations and fact. This may be precisely because the most prominent privacy threats, whether they be Big Data analytics drawing conclusions from aggregating for an unforeseen business end, a network of companies engaged in a wide variety of secondary uses of data shared between them, or an illicit dark web of hackers and fraudsters, are beyond their comprehension.

The second upshot of CI is that it identifies privacy as a property of *information flows*, which when unpacked proves to be a more substantive claim than it may first appear. When we speak about "consumer information" or "personal information", we are faced with the ambiguity of the meaning of the word "information", which can mean alternatively either a medium of representation (such as paper or electronic records, "data") or a mathematical relationship between events or objects such that one is sufficient for inferences about the other [16]. As a physical fact, what we are often concerned with is the exchange and security of data that is identifiable with persons, suggesting the first meaning of "information". This sense of information as a *thing* perhaps encourages privacy solutions that frame personal information as a good that could be protected by property rights

and thereby allocated more efficiently. [13, 21]

However, the sense of information tied to the phrase *information flow* is slightly different. Dretske [7] argues that a message carries information about something it represents if and only if it messages of its kind carry a regular ("nomic", in Dretske's wording) relationship with what is represented. This connects the concept of information flow to the idea in classical information theory [23] of an information *channel* that establishes a correspondence between the probability distributions of two random events.

Subsequent work in statistical probability and causation [18] have complicated the matter by formulating precisely how causally linked events can be correlated without being directly caused by each other. For example, two events that share a common cause can be correlated. When we discuss *information flow*, we may be talking about a causal flow, as when an analyst looks at health data that was originally provided by hospital patients. Alternatively, we may be talking about a more indirect associational flow, as when an analyst uses data about people's addresses and auxiliary information about the environmental condiitons in different urban neighborhoods to draw conclusions about the health of the residents.

Because associational connections between data and persons can occur through so many causal pathways, it may be infeasible to enforce policies based on the associative content of data; rather, policies may be limited to regulating the causal flow of data (cf. [15]. However, since potential secondary data uses depend on associative properties, not causal history, this means there is a mismatch between the power of regulatory policies to manage information flows and the possible uses of data when it is in unregulated hands. This includes the cases when personal data is entrusted to the control of individuals, even when it is *their own personal data.* Since personal data inevitably carries associative information about other persons (because of social homophily, inherited traits, etc.) as well as associations with other information about the same person, there is little reason to suspect that individuals are competent to make rational decisions about the potential consequences of decisions they make. There are key information assymetries between what individuals think they are disclosing and what it is possible to learn from what is disclosed when considering the data in aggregate or in light of other specialized knowledge. This suggests the possibility of market failure.

3 Two-sided Markets

Economic contexts like retail price differentiation, human capacity markets, and expert services each make use of personal information. We have reason to believe that consumers have privacy expectations that are tied to context, and that this makes them ill-equipped to understand the implications of potential secondary uses of their information in contexts other than the one where the information was disclosed. This is despite those secondary uses being potentially significant sources of harm.

An economic context that we have not yet discussed are two-sided markets, where a firm acts as an intermediary between marketers and consumers. This is clearly an important aspect of privacy economics today. It is impossible to ignore the success of the 'Big Five' technology companies– Google/Alphabet, Facebook, Amazon, Apple, and MicroSoft–their impact on the technology business ecosystem in general (cf. [22]), and the fact that two of them (Google and Facebook) are an advertising duopoly [17]. While there are many reasons why these companies are successful, the relationship between personal information, two-sided markets, and industrial organization must be considered in light of prima facie evidence of a natural oligopoly.

A hypothesis is this: while there are many economic contexts (such as those already described) in which personal information is useful for firms, most firms do not have this personal information available to them. However, large two-sided markets that target individuals based on consumer profiles provide a way for smaller firms to pay for the advantage of having this information. It is known, for example, that Google Ads are used to target advertisements for job opportunities [6]. Through the advertising mechanism, firms can test populations for their reservation prices (through ads with 'special deals' on products and services) and entice desireable populations into business (through jobs, credit cards, etc.)

On-line advertising today is able to make use of personal information that might is both very personally revealing and not obviously connected to all the possible contexts which make use of advertising *as an intermediary mechanism*. Behavioral advertising is a widespread practice [9] that in principle makes it possible to use search history to assist retailers in price discrimination. Two-sided markets in a sense provide an omnibus expert service, though they have few if any legal restrictions on what data they collect or how they use it. An important question for consumer privacy is whether two-sided markets have consolidated as a result of market failure.

4 Market failure and regulation

By the preceding argument, consumers are not competent to make decisions about how to control their personal information because their privacy expectations are tied to contexts that are routinely violated in practice. Potential secondary uses of personal data depend on associational properties of the data that are beyond users comprehension. In the case of large, data-rich firms, these associational properties are discovered through aggregation and data mining by the very firms that attract consumer interaction through expert services that they offer. This data is then used in two-sided markets, which act as intermediaries in many other economic contexts, further complicating any prediction of the benefits and harms of disclosure. Quantitative, let alone qualitative, prediction of these harms and benefits is beyond what an individual can accomplish. It is also, incidentally, beyond what independent researchers can reasonably accomplish. Assessing the impact of such market activity requires that one have access to the kind of personal data held by the data-rich firms that are under consideration.

The conditions appear to be ripe for classic market failure, or else there would be if there were a market to begin with. As has been mentioned, there are not property rights for personal data. Personal data is not a good being produced by anybody in the privacy economics ecosystem. It is rather information in the *strategic* sense of allowing some market actors to outperform others. There is no sense in which the market of personal information has the properties that would lead us to believe the market would allocate resources efficiently. Perhaps rather than ask if there is a market failure, we should be asking what is happening, if not a market at all?

As an alternative to regulating personal data as a kind of property, some have proposed regulating personal data through tort [19, 5]. Certainly some meanings of "privacy", such as those that refer to protection from libel, are enforceable in this way. However, as we have discussed it seems unlikely that the scope of consumer harm or benefit can be adequately assessed given the scale of the empirical problems involved.

A third and perhaps more promising alternative is strengthened data protection laws for two-sided markets. As we have noted, in most expert service sectors, including health care, finance, education, and so on, there are existing sectoral data protection laws ensuring confidentiality. The existence of these laws is an indication that without them, these expert service markets would implode in market failure. If protecting confidential information from secondary use (through austere prohibitions on disclosure and secu-

rity investments) is a form of service *quality*, and this quality is difficult for consumers to assess independently, then this information assymetry about service quality would result in a market failure along the lines of Akerlof's market for "lemons". [3] Since unregulated two-sided markets are in the senses described above equivalent to providing unrestricted secondary use to other firms, perhaps present economic conditions are just such a market failure.

The European Union's General Data Protection Regulation is an example of a privacy regulation that might prevent unrestricted secondary use of information. It accomplishes this largely through *purpose restrictions*, which place restrictions on the goals for which collected data may be used. This approach as some resonance with Contextual Integrity [8], though the latter is more sensitive to social expectation and frames privacy in terms of information flows rather than processing of personal data. To the extent that considering personal data to be a *thing* is misleading, perhaps encouraging a property rights based response which would be ineffective, it may more more effective to craft data protection regulation through the framework of dignitary privacy. [20]

5 Acknowledgements

This work draws on collaboration with Michael Tschantz, Helen Nissenbaum, and Anupam Datta. I thank John Chuang as well as Ignacio Cofone, Yafit Lev-Aretz, Helen Nissenbaum, John Nay, Julia Powles, Madelyn R. Sanfilippo, Yan Shvartzshnaider, Katherine Strandberg, Bendert Zevenberger, and other members of the Privacy Research Group at the NYU School of Law for their helpful comments.

I gratefully acknowledge funding support from the U.S. Defense Advanced Research Projects Agency (DARPA) under award FA8750-16-2-0287. The opinions in this paper are those of the author and do not necessarily reflect the opinions of any funding sponsor or the United States Government.

References

[1] Alessandro Acquisti, Curtis R Taylor, and Liad Wagman. The economics of privacy. 2016.

[2] Alessandro Acquisti and Hal R Varian. Conditioning prices on purchase history. *Marketing Science*, 24(3):367–381, 2005.

[3] George A Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. *The quarterly journal of economics*, pages 488–500, 1970.

[4] Ryan Calo. The boundaries of privacy harm. *Ind. LJ*, 86:1131, 2011.

[5] Ignacio N Cofone. The dynamic effect of information privacy law. *Minn. JL Sci. & Tech.*, 18:517, 2017.

[6] Amit Datta, Michael Carl Tschantz, and Anupam Datta. Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, 2015(1):92–112, 2015.

[7] Fred Dretske. Knowledge and the flow of information. 1981.

[8] Mireille Hildebrandt. Slaves to big data. or are we? 2013.

[9] Chris Jay Hoofnagle, Ashkan Soltani, Nathan Good, Dietrich James Wambach, and Mika D Ayenson. Behavioral advertising: the offer you cannot refuse. 2012.

[10] Chris Jay Hoofnagle and Jennifer M Urban. Alan westin's privacy homo economicus. 2014.

[11] Kirsten Martin and Helen Nissenbaum. Measuring privacy: an empirical test using context to expose confounding variables. *Colum. Sci. & Tech. L. Rev.*, 18:176, 2016.

[12] Deirdre K Mulligan, Colin Koopman, and Nick Doty. Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy. *Phil. Trans. R. Soc. A*, 374(2083):20160118, 2016.

[13] Richard S Murphy. Property rights in personal information: An economic defense of privacy. *Geo. LJ*, 84:2381, 1995.

[14] Helen Nissenbaum. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press, 2009.

[15] Helen Nissenbaum, Anupam Datta, Michael Tschantz, and Sebastian Benthall. The need for origin privacy. Technical report, New York University, 2017.

[16] Geoffrey Nunberg. Farewell to the information age. *The future of the book*, pages 103–138, 1996.

[17] Lara O'Reilly. The race is on to challenge google-facebook 'duopoly' in digital advertising, Jun 2017.

[18] Judea Pearl. Probabilistic reasoning in intelligent systems: Networks of plausible inference. 1988.

[19] Richard A Posner. The economics of privacy. *The American economic review*, 71(2):405–409, 1981.

[20] Robert Post. Data privacy and dignitary privacy: Google spain, the right to be forgotten, and the construction of the public sphere. 2017.

[21] Pamela Samuelson. Privacy as intellectual property? *Stanford Law Review*, pages 1125–1173, 2000.

[22] Conor Sen. The 'big five' could destroy the tech ecosystem, Nov 2017.

[23] CE Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[24] Carl Shapiro and Hal R Varian. *Information rules: a strategic guide to the network economy*. Harvard Business Press, 1998.

[25] Yan Shvartzshnaider, Schrasing Tong, Thomas Wies, Paula Kift, Helen Nissenbaum, Lakshminarayanan Subramanian, and Prateek Mittal. Learning privacy expectations by crowdsourcing contextual informational norms. In *Proceedings of the 4th AAAI Conference on Human Computation and Crowdsourcing (HCOMP16)*, 2016.

[26] Daniel J Solove. A taxonomy of privacy. *U. Pa. L. Rev.*, 154:477, 2005.

[27] Hal R Varian. Economics of information technology. *University of California, Berkeley*, 2001.