Article type : Research Dialogue

Corresponding author mail id : jag@umich.edu

Circles of Privacy

H. V. Jagadish,

Director Michigan Institute for Data Science

Bernard A Galler Collegiate Professor of Electrical Engineering and Computer Science

University of Michigan, Ann Arbor, MI

Acquisti et al. have been a leading voice in pointing the ways in which consumers are giving up their privacy, and the consequences that flow from this. See (Acquisti, Brandimarte and Lowenstein, 2015). Others have made similar points, notably Shoshana Zuboff in her influential book on Surveillance Capitalism (Zuboff 2019). This paper by Acquisti, Brandimarte and Loewenstein (ABL) presents an excellent review of the work on psychological biases that make us prone to make poor privacy choices, including both their own work on this topic and that of others. However, there are several important points about privacy that I think are worth stating more forcefully.

Privacy is not binary

There is a long history, in both technology and law, to consider privacy as an all-or-nothing concept. A piece of information is either public or it is private. There are no other possibilities. In practice, this is not how we live. We have what I call ***circles of privacy*** (Jagadish 2015a). There are things you share with your spouse, things you share with your family, things you share with a few close friends, things you share with a wide circle of acquaintances, none of which may be things you'd be willing to make public. In other words, for every private piece of information, you define a *circle* within which you are willing to share it, without divulging outside this circle. This applies not just to friends and family, but also to business relationships: any merchant you transact with may obviously know your credit card number and purchase history, but you probably do not want the merchant to share these with anyone else. Furthermore, who is in this circle may depend on circumstances, and may vary with time. For example, information you were willing to share yesterday with a friend, or a merchant, you may no longer be willing to share today, because you have ended that relationship.

The need for richer definitions of privacy has been eloquently argued for by many. Nissenbaum (Nissenbaum 2009) introduced the notion of contextual integrity for privacy. In every social context, she points out that there are context-relative norms of privacy. Waldman defines privacy as trust (Waldman 2018), her point being that you share information with entities you trust to handle it appropriately. ABL also point out the psychological and economic benefits of sharing information. But this type of sharing is not an abdication of privacy. Rather, it is an integral part of our information management, complementing our privacy needs.

When we share selectively, we have an unwritten "NDA," a non-disclosure agreement. If your friends gossip about something you shared with them in confidence, you will likely end the friendship. What we seek is the ability to limit sharing. If your information is kept only with you, it is easy for you to control. If it is shared, you need mechanisms, such as social norms, regulations, or legal agreements, that prevent undesired additional sharing beyond the circle of privacy you define.

From a technology perspective, it is easy to define a non-disclosure requirement for a computer system. Indeed, we have systems that do a pretty good job of showing you, and only you, information about your own account, and your history, with a merchant or with a health care provider; these systems do not allow you to see such information about others' accounts. [Data breaches do occur from time to time, and are a risk to keep in mind, but at least the expected functioning of the system is clear in normal circumstances, without a breach]. If privacy were a simple binary construct, we could have declared success.

The difficulty we have is that privacy is more than binary. To support this notion, we need to build mechanisms for constrained access, and such nuanced control is technically hard to implement. Once someone has access to a piece of information, it is difficult to use technological means to limit what they can do with it. While it is easy to limit what a user can do within an application, it is not easy to stop a user from copying information out of the application. Once the information has been copied, the user can do whatever they want with it. Copy protection technologies have a long and sorry history (Wallach 2001).

From a legal perspective, there is the "third party doctrine" (Kerr 2008), which holds that you have no reasonable expectation of privacy for information you voluntarily provide to someone else, such as a merchant, or a friend. Once again, there is a binary notion of privacy – no zone that permits limited sharing, no circle of privacy.

In terms of control, there are at least two important aspects. One is control over further sharing: if you disclose information to a friend, can this friend disclose it to others? If so, which ones? If a merchant knows your purchase history, can the merchant share it with other companies that may wish to market to you? The second aspect is re-purposing, which could happen even without additional sharing. You continuously disclose your location to your cell phone company so that it can provide you service, but you do not want it to infer your religion or your favorite stores by analyzing your location trace. A well-known early instance of this mis-step was when Target began to identify customers who may be expecting a baby. This turns

out not too hard to do, based on the purchase of certain items, but was considered creepy by customers, and so was stopped by Target (but not before a teenager had her father find out about a pregnancy through a Target mailer and a New York Times article that was widely noticed) (Duhigg 2012).

Sharing is Not a Choice

If you ask a friend for a ride to the airport, you do not have a choice to keep your trip private, at least from this friend. By social norms and etiquette, you would be expected to tell your friend at least where you are going and for how long. Even if you can avoid getting into the details, your friend will probably be able to guess the nature of your trip based on what you wear, what baggage you are taking, etc.

If you make a purchase from a merchant, the merchant has to know what was purchased and how it was paid for. If you used a credit card or had something shipped to your home, the merchant probably has enough information to know who you are, even if you have not set up an account with the merchant. Based on the third party doctrine, you have voluntarily shared personal information with the merchant, and therefore given the merchant the right to do whatever they wish with it.

Merchants, of course, have little to gain from making your information public: by doing so they only stand to lose your trust. Instead, merchants monetize your information, using it to market to you themselves, as well as selling it to others who can market to you. They may establish privacy policies explaining the ways in which they may do these things. There may even be regulations, e.g. for financial information, that require them to provide you with privacy notification in a particular form and allow you to stop them from certain uses. With these caveats, the third party doctrine still applies. ABL note, correctly, that most of us do not bother to read privacy policies presented to us. I claim this is driven not only by the length of fine

print, but also by lack of recourse: if you wish to obtain service from that provider, you have to accept the policy. If you read the policy and dislike it, you could, in principle, go elsewhere. But, in practice, you expect every other merchant to have a similar policy, and just do not think it worth the effort to comparison shop based on privacy policies.

In this bleak legal landscape for privacy, at least in the United States, a really hopeful sign is a recent decision (Carpenter v. United States) by the US supreme court (United States Supreme Court 2017) that began establishing limits to the third party doctrine. Specifically, it held that the location of a cell phone may be disclosed to a phone company without being considered voluntary disclosure under the third party doctrine; since this disclosure is necessary to obtain phone service, which is an essential service, the disclosure cannot be considered voluntary.

Companies are not people

Humans have evolved to interact with other humans. One should expect that our instincts, and behavioral norms, regarding privacy, have been developed for sharing information with other humans. However, we are now interacting with corporations. So normal human instincts and normal societal norms of behavior may not be appropriate.

For example, the process of two individuals sharing information with each other is somewhat symmetric. This results in natural reciprocity. One reason you can trust your friend not to misuse your private information is because you have some private information from the friend with which you could take revenge. Neither side needs to make a threat, or even to think about doing so consciously. The underlying facts exert influence even in the absence of the slightest threat posture, because they are embedded into social norms.

With a company, the situation is completely asymmetric. The company learns information about the consumer. The consumer usually learns very little about the company beyond what

the company wants to share through its own web site and public relations. There are some exceptions, of course – review sites sometimes provide information about a company or its products that a consumer may find helpful; official filings may occasionally provide useful information too. '

Another crucial difference between companies and people Is that when people die, they "take their secrets to their graves." When companies go out of business, their data assets are monetized by their creditors. Ongoing businesses looking for success in the long run do not want to do things that upset their customers. But a company that has gone out of business no longer has customers it cares about. So, if the use of customer data was limited while the company was in business, that is no longer the case when the company is out of business. In fact, the company's creditors are likely to have suffered significant financial losses due to the company's failure, and will do their best recoup what they can by selling its data, and other assets. I have discussed this problem extensively (Jagadish 2015a). In the U.S., a consumer privacy ombudsperson is appointed to limit this damage (Agin 2012). A bankrupt firm seeking to monetize data assets has to demonstrate to the ombudsperson that the proposed use does not violate assurances explicitly given at the time of data collection.

Information asymmetry

Information is of value, and can benefit any party to a negotiation. As ABL point out, a company can keep some of the consumer's surplus for itself if it has information about the consumer. Information asymmetry can tilt the scales towards the party with more information.

Consumers, relying upon their human abilities, can remember and process only so much information about the merchants they deal with. Companies invest in sophisticated information gathering and analysis systems, and can benefit from this asymmetry.

The more information companies can collect, the greater this asymmetry. Privacy protections can limit what information can be collected and used, and therefore limit this asymmetry. The point here isn't even the specific limits on information – just that having limits reduces asymmetry.

Scale changes everything

People often think about Big Data simply in terms of its size, and the challenges of managing the large volumes. However, there are many implications of Big Data, which we need to give consideration to as I have argued extensively elsewhere (Jagadish 2014, Jagadish 2015b, Jagadish 2016a, Jagadish 2016b). For example, the heterogeneity of Big Data makes it challenging to derive results that go across data sources and data subjects. Similarly, the need for aggregate conclusions compels us to ignore individual nuances, some of which could be material. Perhaps the most important implication is the qualitative impact of scale on the nature of analysis. We are used to small-scale information gathering. Our thinking, and laws, are designed for this small-scale. But things are completely different at large scale, and we need to think about this carefully.

For example, there is a general presumption of no privacy on a public street. If you are walking on the street in your town, whoever sees you can know that you took this walk. You may be observed coming from a place of worship and going into a shop. Most of us are quite okay with this. We are not furtive with our movements. Random passers-by may see us entering a store, but few will recognize us and even fewer will care to remember.

But what if the same observation is continuous and universal, rather than sporadic? Performed by a network of surveillance cameras, we could have every one of our movements recorded and analyzed. While your single visit to a particular store may not be remarkable in any way, the full set of visits to all stores will reveal a great deal about you.

The same principle applies in almost every data collection situation. If I download an article on the web, I expect at least the source website and my internet provider will know that I did so, and I may be comfortable with this. However, a collation of all articles I have read is a different matter altogether. With shopping too, a retailer knowing an item or two I purchased is different from a mega-retailer like Amazon having an overview of my purchases across time and across categories. ABL cite several studies that show an apparently cavalier attitude towards privacy displayed by consumers in many cases. I suspect that some of these studies obtained the results they did because the subjects did not correctly understand the scale and completeness of the surveillance they would be subject to.

Data collection as public good

We have discussed above how information has value, and can lead to financial benefit in business negotiations. It is worth pointing out that information is also of societal value. So, it is not the case that we wish to disallow all collection of information, or even limit it so severely that valuable benefits cannot be obtained.

Perhaps the most salient example of information as a public good is the census. We all reveal considerable personal information to a trusted authority, who then releases this information, in appropriately aggregated form, as a public good. Countless users benefit from this statistical reporting. Of course, the key to making it all work is a carefully designed and well-regulated system that provides individuals with reasonable guarantees of privacy.

Another example of public benefit from private data is with medical research. Health records are sensitive for many of us, and there are good privacy protections for health records by law. Yet, analysis of health data from large numbers of patients has the potential to yield significant new insights, and so is something society should enable in a responsible way.

There are technological solutions available today to address problems of this type, where we can keep private the data of individuals in a set while making it possible to analyze the aggregate. A specific technique called *differential privacy* adds a small amount of noise to computed aggregates, enough to make it very difficult to "reverse engineer" individual records (Dwork 2008).

There are many other uses for data as a public good. For example, consider location tracing, which we normally would consider highly invasive in terms of privacy. Yet, in the context of the COVID-19 pandemic, location history can be vital to contact tracing, and there is robust debate about the extent of privacy encroachment that is appropriate for civil society in the context of a pandemic (Cho 2020).

Privacy is not an individual choice

We have a strong culture of individual choice, and it is easiest to model privacy in terms of an individual's personal information. However, it turns out that we are not all as independent as this simplistic model represents. Consider genetic information. On the one hand, nothing could be more central to an individual than their DNA. On the other hand, we know that the DNA of family members is strongly related. If your sister chooses to make her DNA public, then your DNA is also somewhat public – at least in a probabilistic sense, the public has a pretty good idea of what your DNA may be, even if the information is not perfect. What obligation does your sister have to consult with you before publishing her DNA? How about more distant relatives? They too share DNA with you. A distant cousin's DNA may tell me less about your DNA than your sister's DNA does, but it still gives me quite a bit of information that you consider private. You may perhaps take comfort in the fact that the exact DNA is not completely revealed. But this partial revelation can have consequences as recently demonstrated by the capture of Joseph DeAngelo, a serial killer and rapist in California whose

DNA at the crime scene was entered into a genealogy web site, matched against the DNA of his relatives, leading to two suspects, one of whom was the perpetrator (Guerini 2018).

Associations with other people can occur due to reasons other than family as well. In social networks, there is often similarity of tastes, age, and behaviors between friends. If I know something about several of your friends, I can have a pretty good guess about you as well. In fact, friend information was an important piece of the Cambridge Analytica scandal that ABL opened their article with. ("Many Facebook users whose data was also collected hadn't even interacted with the app.")

Even in the absence of explicit associations, mere sharing of certain attributes can also give away information. Pollsters have long used demographic and geographic attribute values to estimate the likelihood of your supporting a particular candidate. The same idea is used to predict your shopping style and your purchase behavior, based on what others "similar to you" have done.

The bottom line is that information you consider yours may not even be completely yours to give away. In revealing information about yourself, you may be compelling your friends and relatives to reveal at least some information about themselves.

Solutions

Privacy loss is a problem enabled by technology, and one would hope that technology can address this problem. Unfortunately, technology alone will not be enough to provide privacy protection. The discussion above, and more extensively in this dialog, should make it clear that privacy protection will not naturally occur in the free market. Therefore, we are left with the need for regulation as the only way forward, with technology to support both compliance and enforcement. As ABL say, "policy intervention is necessary."

Technology can limit access to data, and this access limiting is the fundamental platform on which to provide privacy protection. However, once access is granted, it typically comes with no restrictions. Technology can also provide some forms of limited access, as in the differential privacy example discussed above. However, these types of access are difficult to define and implement. At present, differential privacy is the only limited access mechanism available with provable guarantees. What we wish is to place limits on sharing and on use. These limits have to be specified by regulation, or contract. Technology can then be used to assist in implementation of the agreed upon behavior. A pure technology solution Is all too easy to defeat.

ABL point out several obstacles that lie in the path to effective regulation. I agree that these are present, and may well lead to ineffective regulation, or worse. Yet, we have other capitalist systems that have developed reasonable regulations. For example, the stock market has many rules, including rules on hard-to-define topics like insider trading. The set of rules is essential in providing confidence to the many players in the market. The rules are far from perfect, and endure constant pressure from influential players who could gain a great deal from their relaxation. Yet, we have a system that most of us find workable. I would like to be optimistic and push towards developing the best regulatory environment we can for privacy, just as we have for the stock market.

In my own field of Computer Science, there was not much discussion of ethics beyond privacy even as recently as five years ago. When I developed my MOOC on Data Science Ethics (Jagadish 2016a and Jagadish 2017), it was in a relative vacuum. But today, the picture is completely different. Computer Science and Data Science educators broadly agree on the need for graduates to have ethics training, and to be able to engage effectively in shaping policy on topics such as privacy. This is a huge shift in mindset, and I am optimistic that it will result in a very different value system informing technical innovation, on issues such as privacy, as these students join the technical workforce and influence decisions made by firms that employ them.

References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509-514.

Agin, W. E. (2010). Reconciling the FTC Act with the Consumer Privacy Ombudsman's Role. American Bankruptcy Institute Journal, 29(8), 38.

Cho, H., Ippolito, D., & Yu, Y. W. (2020). Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs. arXiv preprint arXiv:2003.11511.

Duhigg, Charles (2012). How Companies Learn Your Secrets. *New York Times*, Feb. 16. Available at https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

Dwork, C. (2008, April). Differential privacy: A survey of results. In International conference on theory and applications of models of computation (pp. 1-19). Springer, Berlin, Heidelberg.

Guerrini, C. J., Robinson, J. O., Petersen, D., & McGuire, A. L. (2018). Should police have access to genetic genealogy databases? Capturing the Golden State Killer and other criminals using a controversial new forensic technique. PLoS biology, 16(10), e2006906.

Jagadish, H. V. (2017). Data Science Ethics. Coursera Online Course (MOOC), University of Michigan.

Jagadish, H. V. (2016a). Data Science Ethics. EdX Online Course (MOOC), University of Michigan.

Jagadish, H. V. (2016b). The values challenge for Big Data. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* (2016): 77-84.

Jagadish, H. V. (2015a). Big data and science: Myths and reality. *Big Data Research*, 2(2), 49-52.

Jagadish, H. V. (2015b). Moving past the "Wild West" era for Big Data. *In Proceedings of the 2015 IEEE International Conference on Big Data*. IEEE Computer Society,2. DOI:https://doi.org/10.1109/BigData.2015.7363733

Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, 57(7), 86-94.

Kerr, O. S. (2008). The case for the third-party doctrine. Mich. L. Rev., 107, 561.

Nissenbaum, H. (2009). Privacy in context: Technology, policy, and the integrity of social life. Stanford University Press. ISBN-10: 0804752370

United States Supreme Court (2017). Carpenter v. United States. Available at https://www.oyez.org/cases/2017/16-402

Waldman, A. E. (2018). Privacy as trust: Information privacy for an information age. Cambridge University Press. ISBN-10: 1316636941

Wallach, D. S. (2001). Copy protection technology is doomed. Computer, 34(10), 48-49. https://doi.org/10.1109/2.955098

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power: Barack Obama's Books of 2019. Profile Books. ISBN-10: 1610395697