October 23, 2013
Sloan Cybersecurity Lecture, NYU-Poly
"Reclaim Your Name"

Panel Discussion

Moderator:

Katherine J. Strandburg	Alfred B. Engelberg Professor of Law, New York University

Members:

Julie Brill	Commissioner, FTC
Jennifer Barrett Glasgow	Global Privacy and Public Policy Executive, Acxiom Corporation
Julia Angwin	Technology Journalist, The Wall Street Journal
Daniel Weitzner	Director, CSAIL Decentralized Information Group at MIT

We've heard about the necessity of transparency in data collection and analysis. Should the focus for transparency be on individuals, and how should you give them the information? Or should an additional/alternative approach be directed towards regulators, where disclosure is more like SEC filing instead of being user-directed?

JBG: This is a great opportunity and these are very important issues, but none are easy issues with easy answers. I've been in the privacy profession for a good long while, and the challenges and choices are just getting harder.

We have beat the issue of notices to death. They ought to be for lawyers, not the consumers. We need to figure out a new way to communicate with users about big data, analytics, and how data flows through the organization. We also need to work with how people learn about this. I won't learn from going to websites and learning on my own - I need to be taught about general practices so that then when I am presented with a choice, I have some foundation for making the choice. Nowadays consumers are intimidates so they don't opt - they just do what requires the least energy. Unless they are forced to make a decision, they just don't opt. That speaks to me that they don't understand the choices they are being asked to make. So if you give more choices, they won't be able to make those either.

DW: I want to distinguish transparency from accountability. Transparency is a means, not an end. It makes clear to regulators, to advocates, and to 1/10 of 1 percent of users who read the privacy policy what their policy is. Whatever the FTC has been able to do is based on enforcing deviations from states privacy policies - so transparency is important. But this is not the broader end of privacy.

What we ought to think about as designers is - accountability. What does it mean when data is collected, what is it actually being used for, and can we make it easier for users to understand? Can we detect data misuse?

Our privacy tools are also way behind the big data analytic power - it's your iPhone vs Big Blue and watson.

JA: I've been writing investigative stories about privacy for years and each story is always shocking. But why should we care so deeply about this?

One thing that I found was that before I started to protect my privacy, I did an audit. I made a list of 250 data brokers, found their policies, and was only able to get data from around 30. That is a transparency gap. Once I knew the data was there, I felt a visceral feeling that I just want it. Maybe I can learn new things about myself from this data! It is part of me! People want to see their data.

My axciom data was actually not great. One profile showed me some personality stuff, then another one for 5 bucks gave me my past addresses, and aboutthedata.com said I was a single mother with a 17 year old child. So this was interesting. But boy they were right about my transactions. It was scary. Actually, that led me to realize that maybe i'm doing a little too much online shopping and should cut it out!

JB: Most of my data was incorrect too on aboutthedata.com. Now how does the consumer deal with this? My income was way too low, they show that I barely shop online. So i've been dealing with this for a really long time, and I started to correct it - and then I second guessed myself. But not all consumers realize that this was just marketing data and not necessarily worth correcting.

Some people are highly engaged in this and we depend on them to give the FTC information about where to go look for problems. So the detailed disclosure helps for that. But for the average consumer, those are completely meaningless. To respond to Danny's point, so much more needs to be put under the hood. Too many choices mean they won't opt, which is why we need more privacy by design - allowing some choice but having privacy built in under the hood.

I wanted to follow up on mostly what Danny said, to ask about the question about transparency: transparency as to what? We have very little information about harm or consequences. To what extent are decisions being made based on data? Are there just marketing decisions or other effects? What should we be worried about?

DW: The FCRA focuses attention on the adverse action - denied a loan, job, insurance. Then there is a clear problem for the consumer and there may be a mistake or something wrong. Can we look at ways to extend that model to other adverse actions - maybe you're not being offered purchasing options that you should be offered, or you're getting a different price? There's a technical challenge here where you can see lots of pricing events all over the place. I'd like to know what other Kayak users are paying and what they paid when. There is a lot of transparency involving opening up individual data to other users by giving them the choice. On the other end are nuisances where maybe you get the wrong ad - who cares - but in the middle

is some eligibility information that we ought to know more about.

JA: Can you show the impact? It's hard to trace it all the way through - who knows where your Kayak price came from? So we did a survey to try and find price differences. Staples was customizing prices across the country based on their zip code. That is only the beginning. I worry about the transparency. It took us 9 months to do that study. And what we did violates the Computer Fraud and Abuse Act for sure. I worry about that - we are the watchdogs and we need better tools.

JB: This question of harm is much at the center of the debate in Washington. But we can't crack that nut yet because so many entities are hidden, not consumer facing. We are doing a tremendous amount of enforcement, but also in regard to deception. But a lot more needs to be known. I think there is a harm when health information flows and people don't know about it - do others agree? We are conducting a study of a number of data brokers to see how it flows and is being used, so I hope that will get us further down the road with the harm question.

JBG: i think there is a vehicle here where even the FTC is encouraging industry to step forward and I will criticize my industry that we haven't done enough to establish good norms. We have so many new technologies and we seem to be waiting on laws to catch up to that so we will always be behind by years if not decades. Industry self-regulation is an underused resource.

I want to push on this a little more: to what extent do you think that one way to deal with the harm would be to import some procedural due process where consumers can interact with decisions being made? And ought we to worry about disparate impact on disadvantaged groups? Does that change how we think about how to handle this problem?

JBG: We have a footprint in 17 countries, including Brazil. What if they can't read the privacy policy on their phones - how do they make the choice? Literacy is a challenge! So we really need to think critically and creatively. You can't ask the consumer to figure it all out on their own.

JA: Big data is essentially a way to make decisions based on characteristics. I don't want an experience that discriminates against me. Staples was customizing prices based on how close you live to a competitor's store. That makes sense economically, but that also meant that higher-income people were getting lower prices. And that will end up happening all the time. So big data can actually solidify these realities instead of collapsing these differences.

DW: The SEC model you put forward is fascinating. We need to try to regulate on a scalable basis. The SEC gets miles of paperwork but doesn't really read anything - the gigabytes of data sit there but then reporters read it and stockbrokers read it and lawyers read it to go sue people. That is an example of a scalable regulatory model.

In SEC law, though, we are talking about money, which is very quantifiable. In privacy, we don't

really have a universal metric that we can measure with empirical perspective. There is no currency. Personal data is not the new oil, because it doesn't burn at a certain rate, you can't rate its quality, it doesn't have a numerical price. My colleagues and I have been trying to map legal systems onto computer flows of data - much like how the accounting industry runs, where we can measure stuff with accounting rules to make sure money is going in the right place and where we can detect anomalies.

JB: It's not just data, it's the value of the data - how do we measure it?

JA: But the legal assessments are very dated and the value changes. The history of privacy law has very strict compartmentalization and then other new types of data step in.

DW: Well now there is new data flowing around, so even if the old laws are effective at regulating, then there is a realm of new data with new negative uses and we need to detect and understand them. Law can only evolve based on negative experiences - we can't invent laws without first seeing what creates the problems. We can easily talk about congressional dysfunction right now, but like Jennifer said we can start with industry regulation before law. But again - you can't manage what you can't measure.

JBG: The answers are not simple, and the more data, the more complex the issue becomes to solve. It is easy to isolate one specific question like "do not call" - but then you still have to evolve it over time. Lots of the progress comes more from just talking about the problem than predicting a solution.

And now a question from the audience: Can we talk about the tension between accurate attribution and deidentification?

JB: We at the FTC encourage deidentification as much as possible. And there are many big data research projects that could easily function with robust deID data. There are technological solutions, but the FTC believes that that is not enough - you need to promise that you won't try to reID it and anyone you sell it to needs to agree as well. This is good for medical research, transportation research, etc.

But when it comes to profiling individuals, deID is not useful. Instead, people talk about pseudonymization: we will track a person without their name. However, linkable data is still personally IDable. We picked this up in COPPA (we won't talk about this walled garden right now), where we talk about linkable data.

You need to look at what the purpose of the project is and then determine where you want to go. But so much is focused on individuals, and then deID doesn't really work.

DW: i think there is a very direct tension between deID and these concerns about discrimination. We can make a system where we link the provenance of the outcome, but once it's deID then

it's much harder to track how that outcome happened.

I want to say something more about deID. Once in a while there are these purported silver bullets, usually based on crypto. It feels like magic. But there is a very direct tradeoff between how useful and how identified data is. The census has used this technique for decades to keep identities private to retain trust. But we are in a deID arms race at the moment. Latanya Sweeney was first, and I was on her thesis committee at MIT, and she deidentified the governor's health data. We now have formal characterizations of how good deID is, like differential privacy, but it is still untested technology. We have yet to see it deployed at large scale. It is good for research but probably not for industry.

JBG: we take the approach that deID data today is probably not deID tomorrow. But we need to train people to retain privacy.

JA: Even census data has been used for nefarious purposes - japanese internment camps, identifying suspects post 9-11. And that's just in the US - in other places like Rwanda it is much worse. Humans are flawed and can abuse the data. So consider it radioactive. For example, they had this term called "love intelligence" at the NSA, where people track their spouses.

DW: Most things we have laws about are not susceptible to perfect enforcement. Murder, fraud, ATM machines are all examples. So regardless what we do, bad things will happen. Taht is why we need accountability. We probably would have still had japanese internment and the Holocaust even without census and population lists. So instead of just getting rid of data or regulating, we also need a way to understand and detect misuse of data - because it will happen.

JA: Even though law won't do everything, let's note that law does a lot to mitigate the above problems.

DW: Yes, we need better privacy law in the USA. I acknowledge that.

I also want to put a finer point on the environmental analogy you used above. What is the analogy to personal data? To me, the analogy to radioactivity or carbon emissions is that misuse of data is a byproduct of a process that we think is useful and socially valuable. We don't say stop driving; we say be responsible about it. We have had viable environmental regulations because we can quantify the harm, so we can impose limits.

JB: The problem is we won't have a good measure - for example, something like being profiled as having cancer or being likely to get it. How do you measure that? Especially when it's based on innocuous information and/or purchases.

This is different than pollution because with pollution, the problem is an externality - really big for society at large but not as onerous to the individual. A privacy breach is almost the opposite,

an internality - it may focus on one specific person intently and painfully.

Another audience question: is one viable path an expiration date on collected data?

JB: We hear from marketers that data has a natural expiration date. But yet when we try to talk about time limits there is a lot of pushback.

JBG: I think we run the risk of lumping all data into one bucket. Expiration may help, but it's not necessarily the solution. A date of birth is good data for life! If you're in the market for a new car - that's only helpful for three months. Then again, you can play golf or tennis for 20 or 30 years. Fresh data is helpful but some things are long-term. And some scientific research really needs long-term data like for medicine, etc.

So we have heard three ideas: regulation, industry norms, and ethical training.

JA: Well, in environment issues, we've seen all three collaborating. Think about picking up a dog's waste in a park - thirty years ago, who would imagine that people would actually do it? But laws together with social norms changed that. As far as technological means to protect privacy, though, it's just an arms race - even if you stop using google and encrypt your email, the consumer is very resource-poor as compared to the corporations.

DW: I would emphasize two points from the policy side, though I won't discount the work that tech can do. I think we need more of a baseline of privacy protection that consumers can rely on. We have FCRA, we have HIPAA, but there's a big gap that Jennifer's company works with. Privacy practices are undergoing evolution. I think that transparency, access, and correction rights could be widely implemented and could help us understand what requires more legal attention. Part of what we learned in environmental regulation is that very regimented, command and control-type processes fall behind, so it's better to give companies goals that they must achieve and leave the "how" up to them. Although, some companies are responsible but others are not, like fewer than 10 percent of Android apps that collect personal data even have a privacy policy.

JBG: I think we are in a window right now where we probably won't see legislation until at least 2017. That said, I have a real concern that we will see states enacting their own privacy laws and that could be disastrous. With security laws, we now have different laws in just about every state and now it is just so hard to pass federal law about it. So legally we are in a troubled situation. So I'm urging business circles to be more proactive in developing solutions.

JB: It sounds like we are in violent agreement here. There are lots of debates happening in Washington, Sacramento, and Brussels, and that will continue. Washington may end up following rather than leading. The tech arms race is here, but we do have some companies differentiating themselves on privacy. And it goes both ways, where maybe some companies will take on the consumer cause to help protect consumer privacy.

The point of my talk was to build a bridge, and we need to talk about building more and more bridges. One way is through this ethical training where people like me, Katherine, or Danny can come and talk to the designers and give them some points to think about, such as giving choices to consumers without burdening them and designing privacy for people who are otherwise at a loss about how to choose.