Transparency, Trust, and Consumer Protection in a Complex World
Keynote Address Before Coalition for Networked Information
Fall 2015 Membership Meeting
December 15, 2015

Good afternoon. Thank you, Clifford, for your warm introduction. And thank you to the Coalition for Networked Information (CNI) for inviting me to address your Fall 2015 Membership Meeting. The vast array of topics that you have covered over the past two days is truly inspiring. This scope highlights the profoundly beneficial effects that digital technologies can have on access to information for research, education, cultural preservation, and other scholarly endeavors. I greatly admire the work that CNI has done over the years to connect researchers and educational efforts, and its ongoing efforts to make scholarship more accessible and collaborative.

At first glance, CNI and my agency, the Federal Trade Commission (FTC), would seem to be focused on quite different issues. Yet I think it is clear that CNI and the FTC are wrestling with many of the same questions about the benefits and risks of an increasingly connected and data-driven society. One broad issue in which we share an interest is the Internet of Things. We are connecting nearly everything – from cars and buildings to clothing and light bulbs – to the Internet. Network equipment manufacturer Cisco reports that there are 25 billion networked devices in the world today and predicts that there will be 50 billion by 2020.[1] These sensors, along with our smartphones, tablets, and computers, generate twice as much data today as they did two years ago, and this trend is expected to continue. Sensors that are so small and efficient that they can power themselves with ambient radio waves are becoming a reality.[2] But the number of connected devices and the relentless accumulation of data are only part of the story.

The other part of the story is big data. Data is becoming cheaper to collect and keep, and our ability to analyze it is improving. This development holds many promises. Cities can better maintain their infrastructures by developing sophisticated early warning systems for gas and water leaks. Medical researchers can enroll patients in large-scale research projects and collect streams of useful data that, in the past, would have been a mere trickle coming from surveys and patients' own reports.[3] In classrooms from pre-schools to around them – are making it possible for people all over the world to learn and earn degrees from the world's leading experts and most prestigious institutions.[4]

Some significant risks go along with the potential benefits of connected devices and big data. As we add devices to our homes, classrooms, and clothes, much more sensitive data will be collected. User interfaces on devices will shrink or disappear, making it more difficult for consumers to know when data is being collected, or to exercise any control. In fact, I expect that the Internet itself will soon "disappear" because connectivity will just be part of how things work, as electricity is today.[5]

These developments pose difficult challenges for privacy, security, and fairness in our society. In sensitive settings these challenges are particularly acute, such as education – where privacy violations and security breaches can cause a wide range of harms, and inaccurate or unfair data processing can have big ripple effects in students' lives. More generally, the data that will be available as a result of these connected devices will be deeply personal, and big data analytics will make the data more readily actionable. Some of these devices will handle deeply sensitive information about our health, our homes, and our families. Some will be linked to our financial accounts, some to our email accounts. And devices themselves will be more closely connected with our actions in the physical world, making data security and device security critically important.

But some fundamental aspects of our world will not change, no matter how connected and data-driven we become. Most importantly, we as individuals will remain roughly the same. We will not suddenly become capable of keeping track of dozens or hundreds of streams of our data, peering into the depths of algorithmic decision-making engines, or spotting security flaws in the countless devices and pieces of software that will surround us. Faced with a world of uncertainty about which devices are safe and whether consumers are getting a fair shake in the big data world, consumers could use some help.

I am optimistic that consumers will be able to navigate and benefit from this complex, uncertain, and exciting world. The key, I think, is keeping our focus on the over-arching value of trust. Trust is the flip side of risk. It is the "expectation of favorable reciprocity from others in situations that are uncertain or risky."[6] Trust does not depend on knowing everything there is transparency. And it is this connection between trust and transparency in the context of the Internet of Things and big data that I would like to discuss today.

The FTC's Role in Protecting Consumers' Privacy and Data Security

Before I get to that discussion, let me first describe the role that the FTC plays in privacy, data security, and consumer protection in general. We are the nation's leading consumer protection agency, and we share competition enforcement with the Department of Justice. Under authority given to us in Section 5 of the FTC Act, the FTC is responsible for protecting consumers from a broad range of "unfair or deceptive acts or practices."[8] Under this Section 5 authority, the FTC has brought nearly 100 privacy and data security enforcement actions. The flexibility of Section 5 and our broad authority to obtain remedies that protect consumers have allowed us to keep up with rapid changes in technology. For example, we have brought actions against companies for allegedly collecting information inappropriately from consumers' mobile devices,[9] making unwarranted intrusions into private spaces,[10] exposing health and other sensitive information, exposing previously confidential information about individuals' networks of friends and acquaintances,[11] and providing sensitive information to third parties who in turn victimize consumers.[12] We have also brought hundreds of cases vindicating consumers' rights under laws that protect sensitive information about children,[13] financial information,[14] medical data,[15] and information used to make decisions about consumers' credit, insurance, employment and housing.[16]

The FTC also maintains a busy policy docket. At the beginning of this year, we published a report on the Internet of Things, which emphasizes the importance of data and device security as well as the applicability of established privacy principles to connected devices.[17]

[8] 15 U.S.C. § 45(a).

[9] *See, e.g.*, Goldenshores Techs. LLC C-4466 (F.T.C. Mar. 31, 2014) (decision and order), *available at* https://www.ftc.gov/system/files/documents/cases/140409goldenshoresdo.pdf.

[10] *See* FTC, Press Release, Aaron's Rent-To-Own Chain Settles FTC Charges That It Enabled Computer Spying by Franchisees (Oct. 22, 2013), *available at* https://www.ftc.gov/news-events/press-releases/2013/10/aarons-rent-own-chain-settles-ftc-charges-it-enabled-computer.

[11] *See* Facebook, Inc., C-4365 (F.T.C. July 27, 2012) (decision and order), *available at* https://www.ftc.gov/sites/default/files/documents/cases/2011/10/111024googlebuzzdo.pdf/

[12] FTC v. Sitesearch Corp., d/b/a LeapLab (D. Az. Dec. 23, 2014) (complaint), *available at* http://www.ftc.gov/systems/files/docu aumenTJ/TT6 01s-9(e De03.8()(33(e)-1.o5eyHhOD-.un988 Tm 0 0 1 scn72 270.24 20

Before that, we published a detailed study of the data broker industry,[18] which was in the big data business long before the words "big data" became part of our policy lexicon. We also held public workshops on so-called alternative consumer scores[19] and the potential for big data analytics to be used in ways that discriminate against consumers.[20] And, just last month, we held a public workshop on cross-device tracking, which refers to companies' efforts to correlate a consumer's activities as she moves from smartphone to tablet to desktop computer.[21]

Transparency, Privacy, and Trust

So let me now turn to the connection between prefer[] Spanish").[25] These practices were recently brought to light by the FTC and others.[26] While consumers might benefit from some segmentation – by receiving more relevant advertising, for example – consumers should have choices about where their data ends up and how it is used.

Transparency is enhanced by giving consumers "just in time" information, at key moments when it is most relevant to them, such as when they are deciding to download an app or make purchases on a connected device. But transparency should also include helping consumers navigate the complex ecosystem of data, devices, and big data analytics operating behind the scenes, so that consumers understand the practices that can affect them, and exercise choices about the practices.

Think about the alternatives to being transparent with consumers. One reasonable response of consumers is that they will harbor suspicions about a product or service and may choose to avoid it, or use it less than they would if they trusted it fully. Another response – particularly when an entity that consumers don't know about is collecting or using data – is to react angrily when the truth about the company's data practices come out. And I think it's wise to presume that the truth will come out eventually. In either case, the result is the same: Consumers or customers lose trust in a company, and the results for the company can be devastating.

Many companies and organizations understand this important connection between transparency, privacy, and trust. But being transparent in this data-intensive age is challenging. With the Internet of Things, many connected devices do not have a user interface to present information to consumers about data collection. Devices are becoming more numerous, adding to the mountain of information that companies present to consumers in privacy policies. As devices become integrated into homes and other physical spaces, there are also questions around who should receive disclosures about data collection and use practices. How will the consumers who buy a device – and the innocent bystanders around them – know when a device is recording images or audio? And there are other questions, like how can consumers choose to avoid having their data collected? For how long will their data be kept by the companies who are collecting it? And how will these companies keep the data secure?

Companies that provide connected devices should recognize that providing transparency will require some creative thinking. Visual and auditory cues, and immersive apps and websites should be employed to describe to consumers, in meaningful and relatively simple way, the nature of the information being collected. The same signals should be used to provide

[25] *Id.* at 20 n.52.

[26] *See, e.g.*, CBS News, 60 Minutes; The Data Brokers: Selling Your Personal Information (last updated Aug. 24, 2014), *available at* http://www.cbsnews.com/news/data-brokers-selling-personal-information-60-minutes/; World Privacy Forum, *The Scoring of America* (2014), *available at* http://www.worldprivacyforum.org/2014/04/wpf-report-the-scoring-of-america-how-secret-consumer-scores-threaten-your-privacy-and-your-future/; U.S. Senate Committee on Commerce, Science, and Transportation, *A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes* (Dec. 2013) (staff report), *available at* http://www.commerce.senate.gov/public/?a=Files.Serve&File_id=0D2B3642-6221-4888-A631-08F2F255B577.

consumers with choices about whether any of this information can be used by entities or persons who fall outside the context in which the consumer is employing the device, and in which the consumer expects her information to remain private.

Another promising tool for providing information to consumers, as well as allowing them to exercise meaningful choices, is the "command center" that companies are now developing to run multiple household connected devices.[27] The driving force here is convenience, but these command centers could also provide an opportunity for consumers to understand the information their devices are generating, and to control where that information goes. After all, if you can have a centralized interface to program your garage door, thermostat, television, refrigerator, and who knows what else, you ought to be able to use that same interface to make meaningful choices about the data your devices will collect and where they will send it.

Transparency, Fairness, and Trust

Now let me turn to the issue of transparency and fairness in big data analytics and how it relates to consumer trust. Big data analytics, employed both by data brokers and within companies themselves, are being used to segment consumers by interests and traits and to make an increasingly wide range of decisions about them. Some of these analytics projects could create questions about fairness. For example, a company might analyze its own data in an effort to identify "good" versus "troublesome" customers. But what if this analysis ends up tracking individuals along racial or ethnic lines. A *Harvard Business Review* article argues that this kind of result isn't just possible, but inevitable.[28]

Transparency for big data analytics – both within companies and through third parties –is necessary to engender trust by informing consumers about the significant impact that big data analytics can have on them and clarifying their choices with respect to some kinds of data collection and use. Transparency is also a helpful check on potentially troublesome data practices, since as Louis Brandeis famously said, "sunlight is said to be the best of disinfectants."[29] Without more transparency in big data analytics, questions will linger about the role that big data analytics plays in the marketplace and whether consumers are being treated fairly. But the question is how to present information that is meaningful to consumers.

Many of these kinds of decisions are based on some kind of score – a number generated by an algorithm that gives some indication of what a consumer is likely to be interested in or how she is likely to behave. A familiar example is the credit score. Credit scores are basically predictions of how likely a consumer is to repay a debt. The higher the score, the lower the credit risk. In their early days, credit scores were used strictly for credit decisions – whether you

[27] *See* Don Clark, *The Race to Build Command Centers for Smart Homes*, WALL ST. J. (Jan. 5, 2015), *available at* http://www.wsj.com/articles/the-race-to-build-command-centers-for-smart-homes-1420399511.

[28] *See* Michael Schrage, *Big Data's Dangerous New Era of Discrimination*, HARVARD BUSINESS REVIEW BLOG NETWORK (Jan. 29, 2014, 8:00 a.m.), *available at* http://blogs.hbr.org/2014/01/big-datas-dangerous-new-era-of-discrimination/.

[29] Brandeis Univ., Louis D. Brandeis Legacy Fund for Social Justice, *available at* http://www.brandeis.edu/legacyfund/bio.html (last visited Dec. 14, 2015).

would qualify for a mortgage, for example, and the interest rate that you would be offered. Over time, the use of these same credit scores expanded to other major decisions about consumers, such as whether a prospective employer would extend a job offer to an applicant, or an insurance company would charge a higher premium on auto or homeowners insurance.

We know a great deal about what information is in our credit reports and what our traditional credit scores are for a simple reason. Congress has required some transparency in credit scoring. In 2003, Congress instructed the FTC and the Federal Reserve to study whether one type of popular credit score used for auto insurance employed factors that serve as proxies for race, gender, or other traits that could give rise to unlawful discrimination.[30] In addition, Congress required credit bureaus to make consumers' credit reports available to them for free,[31] and credit scores are increasingly becoming available for free to consumers.[32]

These transparency requirements and practices have been good for consumers, credit bureaus, and companies that rely on credit scores to make business decisions.[33] With the information that goes beyond traditional credit files, to score consumers raises fresh questions about whether these alternative scores may have disparate impacts along racial, ethnic, or other lines that the law protects.

And an increasing range of algorithmic scores and decisions fall outside of the framework provided by the Fair Credit Reporting Act. The FTC identified a few of them in our May 2014 report on data brokers. We highlighted so-called "risk mitigation" services as sources of potentially significant decisions about consumers that are not subject to the specific protections of the FCRA.[36] These services answer questions like "Is this consumer who she claims to be?" and "Is the purchase that this consumer is attempting to make likely to be fraudulent?" While some uses of these "risk mitigation" scores may fall under the FCRA, an important set of them does not.

Transparency is important across the full range of decisions that I've illustrated. But it's not realistic to rely on the approach that the FTC took to understand one type of score used for auto insurance to gain an understanding of the full spectrum of scoring models used today. It took the FTC nearly four years to conduct its study. The FTC – and all other federal agencies for that matter – simply do not have the capacity to study every score out there. This approach simply will not scale.

Moreover, scoring algorithms and other forms of big data analytics rely on statistical models and data system designs that few on the outside understand in detail. And even if we on the outside could peer into the hundreds of scoring algorithms that could potentially affect consumers, what would we learn? We might learn which features of a data set are used in a given algorithm, and what weight a company attaches to them. These details might be so abstract, and so rapidly changing, that they do not tell government, consumers or other concerned stakeholders much at all about what really matters – which is how the algorithms are actually used and whether they have discriminatory or other inappropriate effects.

This suggests that testing the *effects* of big data analytics may be a promising way to go. This route, too, presents some challenges. On a technical level, many companies will not have data to answer definitively the question of whether they are treating consumers of different races or ethnicities differently. For example, an ad network might track consumers by using an email address or a device identifier. The ad network might be able to combine this information with other data that is readily available to it, such as which apps a consumer uses, to make inferences about more sensitive personal characteristics. But this is far different from having test data in which consumers' race, health conditions, financial status, or other sensitive personal characteristics are known.

Doing this kind of analysis from the outside is difficult. Researchers have done some proof-of-concept studies, but they required considerable work and involved efforts to tackle some cutting-edge research questions.[37]

[36] *See* DATA BROKER REPORT, *supra* note 18, at 32-34.

[37] *See, e.g.*, Amit Datta, Michael Carl Tschantz, and Anupam Datta, *Automated Experiments on Ad Settings: A Tale of Opacity, Choice, and Discrimination, in* PROC. ON PRIVACY ENHANCING TECHS. (PETS) 2015, *available at* http://www.andrew.cmu.edu/user/danupam/dtd-pets15.pdf; Latanya Sweeney, *Discrimination in Online Ad*

This means that companies using scoring models should themselves do more to determine whether their own data analytics result in unfair, unethical, or discriminatory effects on consumers. In addition to scrutinizing their own practices, companies can provide consumers with creative UIs to give consumers more meaningful, usable access to their data.

Ultimately, I believe we need legislation to address many of these issues. Technologists have a key role to play, too. They have the skills to make data access tools that are easy for consumers to use, and they have the technical insights that are necessary to determine whether specific analytics practices pose risks of excluding, or otherwise placing at a disadvantage, groups defined according to sensitive traits. I am hopeful that companies will give technologists, including designers and user interface experts, the support and resources needed to tackle these critically important challenges.

* * * * * *

Although I have addressed transparency, trust, privacy, and fairness from the consumer perspective, I hope that the challenges of the Internet of Things a.6(ai.7db stlentv tra3-1ydN6tooleIye)-.6sugj1tb