Commissioner Julie Brill
"Big Data and Consumer Privacy:
Identifying Challenges, Finding Solutions"
Address at the Woodrow Wilson School of
Public and International Affairs
Princeton, University
February 20, 2014

Thank you, Ed, for that kind introduction and for inviting me to speak today. It is always a pleasure to come home to Princeton, and today is no exception. Princeton, and in particular the Woodrow Wilson School, cultivates leaders of all types and in many fields, including those that have helped fuel our global technological revolution. As a lifelon thb1(f4(s)-1hom)-2(i)-4(r)3podr tectiod acae, echnol(g)10(i)1-2(e)4(s)-1n. Ts10(oda)-16(y)20, r

[1] And every time we go online or use a smartphone or credit card, our purchases and movements are tracked.

In a real sense, we are becoming the sum of our digital parts.

The estimates of the data we collectively generate are staggering. One estimate, already more than two years out of date, suggests that 1.8 trillion gigabytes of data were created in the year 2011 alone – that's the equivalent of every U.S. citizen writing three tweets per minute for almost 27,000 years.[2] Ninety percent of the world's data, from the beginning of time until now,

[1] *See* Lisa Wirthman, *What Your Cellphone Is Telling Retailers About You*, FORBES EMCVOICE, Dec. 16, 2013, *available at*

This examination is becoming all the more urgent as phones, cars, and other everyday objects join the Internet of Things. Again, the potential benefits may be profound. Medical wearable devices—such as Google's contact lenses that help diabetics track glucose levels in their tears[11]—have the potential to affect millions of people suffering from a wide range of health conditions. But "smart" devices are about to become always-on sources of deeply personal information. This will be a big shift for consumers. Instead of having a handful of devices – a smartphone, tablet and laptop – that mainly serve to connect consumers to the Internet, consumers may have many devices that they buy for one purpose – making coffee, storing food, driving to work – but that collect and use a vast amount of personal information about them. Whether it is a connected car, home appliance, or wearable device, the data that these connected devices generate could be higher in accuracy, quantity, and sensitivity, and – if combined with other online and offline data – could have the potential to create alarmingly personal consumer profiles.

Will consumers know that connected devices are capable of tracking them in new ways, especially when many of these devices have no user interface? Will companies that for decades have manufactured appliances and other "dumb" devices take the steps necessary to keep secure the vast amounts of personal information that their newly smart devices will generate? And how will the new data from all of these connected devices flow into the huge constellation of personal data that already exists about each of us?[12]

These questions echo the ones that have long surrounded the vast amount of data collection and profiling performed by ad networks, data brokers, and other entities that consumers generally know nothing about because they are not consumer facing. In some instances, these entities track consumers' online behavior. In other instances, these entities merge vast amounts of online and offline information about individuals, turn this information into profiles, and market this information for purposes that may fall outside of the scope of our current regulatory regime.

As we further examine the privacy implications of big data analytics, I believe one of the most troubling practices that we need to address is the collection and use of data — whether generated online or offline — to make sensitive predictions about consumers, such as those involving their sexual orientation, health conditions, financial condition, and race.

Let's look at a well-known, and by now infamous, example. Before Target made news for a data security breach that may involve 110 million consumers' credit cards and debit cards,

the comFBd /TT1 1-1fwi6d [(n)-49-36.1.651, uarh584a-2(r-4(92 Tw [(/TT3 1-1fw)mp)6(lic)6(a)3-10(r)Y)86

Start: Young Single Parents," and "Credit Crunched: City Families."[19] I am concerned that the names and descriptions of such products likely appeal to purveyors of payday loans and other financially risky products to help them identify vulnerable consumers most likely to need quick cash.[20]

Some argue that if data brokers aren't employing predictions about health conditions or other sensitive personal traits for legally forbidden uses, then what is the harm? These advocates will say that predicting consumers' health conditions could help them reduce their risk of disease or make them aware of new opportunities for clinical trials, and predicting their financial situation could help them find new opportunities for credit – benefits that outweigh any breach of privacy. But this view fails to account for the growing level of concern that consumers have about their most sensitive information being collected and stored in individual profiles and used for purposes that consumers do not know about and therefore cannot control.

I believe we should all be concerned about the use of deeply sensitive personal information to make decisions about consumers, outside a legal regime that would provide notice and an opportunity to challenge the accuracy of the data. Similarly, we should be concerned about the risk that such sensitive personal information may fall into the wrong hands through a data breach. But more fundamentally, I believe we should be concerned about the damage that is done to our sense of privacy and autonomy in a society in which information about some of the most sensitive aspects of our lives is available for analysts to examine without our knowledge or consent, and for anyone to buy if they are willing to pay the going price.

These concerns, of course, are not limited to the world of commercial data brokers. We don't have to pass judgment on the revelations about the NSA and other intelligence agencies' data collection and use practices to acknowledge that the recent disclosures have sparked a necessary and overdue debate on how to balance national security against citizens' privacy rights. For those of us who have been looking at the issue of privacy in the Internet age for several years, there is a further benefit: Americans are now more aware than ever of how much their personal data is free-floating in cyberspace, ripe for any data miner – government or otherwise – to collect, package, use – and on the commercial side – sell.

But with that knowledge comes power – the power to review, this time with eyes wide open, what privacy means – or should mean – in the age of the Internet. I believe that's what President Obama meant in June, and again last month, when he noted that the "challenges to our privacy do not come from government alone. Corporations of all shapes and sizes track what you buy, store and analyze

for a "national conversation…about…the general problem of … big data sets, because this is not going to be restricted to government entities."[22]

During our ongoing discussion about government surveillance, national security, and privacy, leaders within the business community have joined the President in recognizing that rebuilding the trust of individuals is essential to the success of all programs and services – both governmental and commercial – built on big data analytics.[23] These business leaders have urged companies to adopt enhanced privacy protections as a key part of strengthening consumer trust.

I agree. While I firmly believe that the national security issues must be addressed separately from the commercial privacy issues, I also firmly believe that the promise of big data – the huge benefits that we may reap from appropriately tailored use of big data analytics – will not be reached until society addresses some of the key consumer privacy concerns stemming from the creation, collection and use of sensitive consumer data and profiles.

I'd like to highlight four steps that I believe should be– -6(t b)2(e)6d3( )]TJ [(0.004 Tw)-1(e)4(oJ 0 T

deidentification technologies and social agreements to not reassociate deidentified data with particular individuals. This means that companies should do everything technically possible to strip their data of identifying markers; they should make a public commitment not to try to re-identify the data; and they should contractually prohibit downstream recipients from doing the same.[25]

Robust deidentification efforts along these lines will solve some of the problem. And the creation of more effective tools to deidentify data – something that I am confident that many in this room could develop – would also help. But such robust deidentification will not solve the problem of big data profiling. The entire data broker enterprise seeks to develop greater insight into the activities, status, beliefs, and preferences of *individuals*. The data the industry employs are therefore about or linkable to individuals.[26]

2. Create Institutional Ethical Monitoring

Another solution offered to the challenges big data presents to privacy is the creation of entities that monitor the ethical use of data. One proposal calls for the creation of "Consumer Subject Review Boards" to determine whether particular projects using consumer data are both legal and ethical.[27] Another proposal calls for individual companies to appoint "algorithmists" – licensed professionals who would have ethical responsibilities for an organization's appropriate handling of consumer data.[28] And I know that Ed Felten's students in Princeton computer science and engineering programs are encouraged to examine the ethical implications of designing algorithms, computer programs, and other innovative projects. More engineering and computer science schools should follow Professor Felten's lead. Yet ethically trained computer scientists, algorithmists, and Consumer Subject Review Boards will only thrive in firms that thoroughly embrace "privacy by design" – from the engineers and programmers all the way up to the C-suite – firms that understand the legal and ethical dimensions of the use of algorithms to make decisions about individuals.

---

[25] *See* FED TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAEKRS 21 (2012), *available at* http://www.ftc.gov/os/2012/03/120326privacyreport.pdf.

[26] *See* John Deighton & Peter Johnson, *The Value of Data: Consequences for Insight, Innovation & Efficiency in the U.S. Economy* (DMA Data-Driven Marketing Institute, Oct. 8, 2013), *available at* http://ddminstitute.thedma.org/#valueofdata.

3.

Policy makers, academics, consumer advocates, and business leaders are all encouraging industry to take more aggressive action to protect consumer privacy. C