Big data will have important, even transformative uses. But consumers, policy makers, and academics also see threats from these vast storehouses of data. This00012u-3 n.Cc --0.002 Tc 0-ba3 0 Td[(

still obey privacy principles that protect consumers. I usually talk about these issues with industry leaders or policymakers in Washington, and I advocate for legal regimes and industry best practices that improve consumer privacy protections. But I've come to realize that we need more than law and more than "best practices" to safeguard privacy effectively. We also need new technological solutions to enhance consumer privacy.

Which brings me to you. Many of you are engineering students and professors, company chief technology officers, and computer scientists. This is your technological revolution. But you understand that technology brings challenges too and I believe that you are passionate about finding solutions. Policymakers like me and my FTC colleagues need to work hand-in-hand with you in the engineering and scientific communities. This is your "call to arms" – or perhaps, given who you are, your "call to keyboard" – to help create technological solutions to some of the most vexing privacy problems presented by big data.

Of course, you won't be like Gary Cooper in *High Noon*, fighting the outlaws all on your own. The world of big data is not quite the Wild West. We have important rules in place governing the ways certain kinds of data can be used. One is the Fair Credit Reporting Act, or "FCRA." And it is the FCRA that presents the first set of challenges for technologists to address.

**First Challenge: The Fair Credit Reporting Act**

The FCRA was our nation's first "big data" law. The seeds for it were planted in the aftermath of World War II. As the economy began to grow, businesses formed cooperatives to enable quicker and more accurate decisions about creditworthiness by sharing information about consumers who were in default or delinquent on loans.[11] Over time, these agencies combined, paving the way for consumers to gain access to credit, insurance and jobs. As credit bureaus increased their ability to draw inferences and make correlations through ever-larger databases, unease about the amount of information that credit bureaus held – as well as its accuracy and its use – also increased. To respond to these concerns, in 1970 Congress passed the Fair Credit Reporting Act.

FCRA governs the use of information to make decisions about consumer credit, insurance, employment and housing. Entities collecting information from multiple sources and selling it to companies making these important decisions must ensure the information is as accurate as possible and used only for approved purposes. Not only does FCRA regulate the use of consumer data, it also gives consumers important rights: Consumers are entitled to access their data, challenge its accuracy, and be notified when they are denied credit or get a loan at less than favorable rates because of negative information in their files.

[11] *See* Mark Furletti, *An Overview and History of Credit Reporting* (Payment Cards Center, Federal Reserve Bank of Philadelphia, June 2002), at 3-4.

My agency, the Federal Trade Commission, enforces the FCRA. Of course, we bring enforcement actions against traditional credit bureaus.[12] Increasingly, though, we are focusing on data brokers that collect information about consumers from offline and online sources, including social media, and then develop and sell apps and other online services for employment and tenant screening, criminal background checks, and other activities plainly covered by FCRA.[13]

FCRA is not a panacea. The process of collecting data, and synthesizing that data into profiles relating to individual consumers, is too error-prone for too many Americans.[14] The FTC's study of the accuracy of credit reports found that the reports for one in twenty U.S. consumers – 10 million people – had serious errors that could result in them receiving less favorable credit than they deserve. We all know it can be a long, arduous and extremely exasperating effort to correct a faulty credit report. Consider

through pictures, graphs or other simple terms the data the device collects about consumers, the uses of the data, and who else might see the data. The smartphones and tablets we all carry – and soon smart watches and connected glasses – create a ready canvas for immersive apps that will provide a new way of giving notice and consent that is more meaningful and less confusing for users.

**Third Challenge: Increased Transparency Mechanisms**

I've saved the toughest challenge for last: the vast amount of data collection and profiling that occurs by entities that are not consumer facing. These entities, called data brokers, merge vast amounts of online and offline information about consumers, turn this information into profiles, and market this information for purposes that may fall outside the FCRA. There are three categories of data brokers' practices worth focusing on.

First, there are those who are selling consumer-specific data for purposes that fall right on – or just beyond – the boundaries of FCRA and other laws. Take for example the new-fangled lending institutions that forgo traditional credit reports in favor of their own big-data-driven analyses culled from social networks and other online sources.[19] Or consider the eBureau, a company that prepares rankings of potential customers based on their "occupation, salary and home value to spending on luxury goods or pet food, … with algorithms that their creators say accurately predict spending."[20] These "e-scores" are sold to determine the customers that are worth wooing on the web.[21]

It can be argued that e-scores don't yet fall under FCRA because they are used for marketing and not for determinations on ultimate eligibility. But when financial institutions – banks, credit and debit card providers, insurers – send targeted ads to targeted consumers advertising certain rates that the institution would be willing to give the consumer based on the e-score, a consumer may never know that she is eligible for an even better rate. These ads are certainly the first cousin, if not closer kin, of firm offers of credit governed by the FCRA. Yet without FCRA protections, a consumer would not know if her e-score led to a higher loan rate or insurance premium, nor would she be able to access and correct any erroneous information about her.

In a hospital or medical trial, or some other context where our federal health privacy law, known as HIPAA,[25] applies, this argument has some force. But when health information flows outside the protected HIPAA environment, I worry about three things. First, how sensitive health information can be used to make decisions about eligibility that fall outside the contours of the FCRA, without notice to the consumer or an opportunity to challenge the accuracy of the data used to make the decisions. Second, what happens if sensitive health information falls into the wrong hands through a data breach? And third, what damage 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?

One way to solve this problem is to ensure that the health data are truly deidentified. Of course, merely stripping identifiers such as names and addresses is not sufficient; it is too easy to re-identify data. But a standard that requires companies to make it impossible to re-identify data could make it effectively useless. FTC has developed best practices around deidentification[26] that strike an appropriate balance by requiring companies to employ reasonable efforts to de-identify data, to publicly commit to use that data only in their de-identified form, and to impose legal requirements to make sure any downstream recipients of de-identified data agree not to re-identify them. And you can join the corps of computer scientists that continue to upgrade deidentification techniques.[27]

But more robust deidentification will not solve the problem of big data profiling. The entire data broker enterprise is aimed at developing greater insight into the activities, status, beliefs, and preference of *individuals*. The data the industry employs are therefore about or linkable to individuals – or as one of the industry's trade associations just-released report refers to it – "individual-level consumer data".[28]

Another solution offered to the challenges big data presents to privacy is the creation of the "algorithmist" – a licensed professional with ethical responsibilities for an organization's appropriate handling of consumer data.[29] But the algorithmist will only thrive in a firm that thoroughly embraces "privacy by design," from the engineers and programmers all the way up to the C-suite, and understands that the use of algorithms to make decisions about individuals has legal and ethical dimensions. NYU Poly and other top notch engineering and computer science schools cover ethics in their courses, but the schools and profession should require more

---

[25] In the United States, the Health Insurance Portability and Accountability Act of 1996 (HIPAA), 42 U.S.C. § 1320d-9, requires hospitals, doctors, health insurance companies and their business partners are required to follow strict guidelines on how they handle health information about patients and insureds. And Institutional Review Boards ensure that human research is conducted ethically, including by maintaining the privacy of research subjects. *See* Dep't of Health and Human Svcs., Office of Human Research Protections, *Institutional Review Board Guide Book* at Chapter 3 (last updated 1993), *available at* http://www. hhs.gov/ohrp/archive/irb/irb_guidebook.htm.

[26] *See* FTC, 2012 PRIVACY REPORT, at 21.

[27] *See, e.g.*, Cynthia Dwork, *A Firm Foundation for Private Data Analysis*, 54 COMMS. OF THE ACM 86-95 (2011), *available at* http://research.microsoft.com/pubs/116123/dwork_cacm.pdf, and references cited therein.

[28] 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/files/2013/10/The-Value-of-Data-Consequences-for-Insight-Innovation-and-Efficiency-in-the-US-Economy.pdf .

[29] MAYER-SCHÖNBERGER & CUKIER, *supra* note 9, at 180 – 182 (2013).

systematic ethical training for undergraduate and graduate degrees. Law schools do this, and you don't ever want to be accused of lagging behind lawyers in terms of ethical training!

Unfortunately, even if industry embraces privacy by design and we license all of you as a new cadre of algorithmists, we will not have met the fundamental challenge of big data in the marketplace: that is, consumers' loss of control of their most private and sensitive information.

Changing the law would help. When I talk about these issues in Washington, I call on Congress to enact legislation that would require data brokers to provide notice, access, and correction rights to consumers scaled to the sensitivity and use of the data at issue. Such a law should require data brokers to give consumers the ability to access their information and correct it when it is used for eligibility determinations, and the ability to opt-out of information used for marketing.

But together we can begin to address consumers' loss of control over their most private and sensitive information even before legislation is enacted. I suggest we need a comprehensive

by seeing certain categories of information, like personal characteristics, home, vehicles, household finances (including credit), purchases, and interests.

Consumers can correct this information. And importantly, consumers can suppress any of the data they see. This is a valuable option; if you don't want to correct erroneous data, or you simply don't want things like your income, race, or marital status to be used in your marketing profile, you can tell Acxiom to stop using it. Consumers can also opt out of Acxiom's marketing profile system altogether.

But there is still more work to do. Acxiom's site provides some transparency, but does it show consumers all the marketing information that's relevant? One reviewer reported that the current site "leaves out many data elements that Acxiom markets to its corporate clients."[31] Moreover, though the option to suppress data is valuable, consumers would have trouble finding it. Allowing consumers to suppress data more easily would be a welcome improvement. Consumers also should not mistake suppression or an opt-out for deletion or the end of data collection. Although Acxiom will not use suppressed data for marketing purposes, the data will stay put. Perhaps most importantly, Acxiom's site currently only shows consumers their data used for marketing purposes. Acxiom holds many other data sets used for eligibility and other key decisions about consumers. Acxiom should take similar steps to provide more transparency about these data sets as well.

Still, I believe Acxiom is on the right road. And you can work with Acxiom to bring it farther down this road, and with other data brokers to help them take the first necessary steps. And then, you can develop an industry-wide, one-stop shop to enable consumers to easily find out who the major data brokers are, and what choices they offer with respect to access, suppression and correction of their data.

My "call to arms" to technologists is not meant as an abdication of the responsibility that law enforcement, policy makers, Congress, industry and other stakeholders have to address these issues. We all have a vital role to play. But it is important to recognize that you – the computer scientists, the engineers, the programmers, the technologists – have a unique set of skills that are key to solving these critical privacy issues. If you join me in this effort, I think that together we can help big data operate in a system that respects consumer privacy and engenders consumer trust, allowing big data to reach its full potential to benefit us all.

Thank you.

[31] *See id.*; *see also* Bant Breen, *Misadventures in Transparency: Data Site Comes Up Short*, DIGIDAY (Sept. 30, 2013), *available at* http://digiday.com/platforms/misadventures-in-transparency-data-site-comes-up-short/.