NBER WORKING PAPER SERIES

ARTIFICIAL INTELLIGENCE AND CONSUMER PRIVACY

Ginger Zhe Jin

Working Paper 24253
http://www.nber.org/papers/w24253

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2018

I am grateful to Ajay Agrawal, Joshua Gans, and Avi Goldfarb for inviting me to contribute to the 2017 NBER Economics of Artificial Intelligence Conference, and to Catherine Tucker, Andrew Stivers and the conference participants for inspiring discussion and comments. All errors are mine. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Artificial Intelligence and Consumer Privacy
Ginger Zhe Jin
NBER Working Paper No. 24253
January 2018
JEL No. D04,D18,D8,L15,L51

ABSTRACT

Thanks to big data, artificial intelligence (AI) has spurred exciting innovations. In the meantime, AI and big data are reshaping the risk in consumer privacy and data security. In this essay, I first define the nature of the problem and then present a few facts about the ongoing risk. The bulk of the essay describes how the U.S. market copes with the risk in current policy environment. It concludes with key challenges facing researchers and policy makers.

Ginger Zhe Jin
University of Maryland
Department of Economics
3115F Tydings Hall
College Park, MD 20742-7211
and NBER
jin@econ.umd.edu

Artificial Intelligence and Consumer Privacy

Ginger Zhe Jin
University of Maryland & NBER
December 18, 2017

Thanks to big data, artificial intelligence (AI) has spurred exciting innovations. In the meantime, AI and big data are reshaping the risk in consumer privacy and data security. In this essay, I first define the nature of the problem and then present a few facts about the ongoing risk. The bulk of the essay describes how the U.S. market copes with the risk in current policy environment. It concludes with key challenges facing researchers and policy makers.

1. Nature of the problem

In early 1980s, economists tended to think of consumer privacy as an information asymmetry *within* a focal transaction: for example, consumers want to hide their willingness to pay just as firms want to hide their real marginal cost; and buyers with less favorable information (say a low credit score) prefer to withhold it just as sellers want to conceal poor product quality (Posner 1981; Stigler 1980). Information economics suggests that both buyers and sellers have an incentive to hide or reveal private information, and these incentives are crucial for market efficiency. In the context of a single transaction, less privacy is not necessarily bad for economic efficiency. Data technology that reveals consumer type could facilitate a better match between product and consumer type; and data technology that helps buyers to assess product quality could encourage high quality production.

New concerns arise because technological advances, which have enabled radical decline in the cost of collecting, storing, processing and using data in mass quantities, extend information asymmetry *far beyond* a single transaction. These advances are often summarized by the terms "big data" and "AI". By big data, I mean large volume of transaction-level data that could identify individual consumers by itself or in combination with other datasets. The most popular AI algorithms take big data as an input in order to understand, predict and influence consumer behavior. Modern AI, used by legitimate companies, could improve management efficiency, motivate innovations, and better match demand and supply. But AI in the wrong hands also allows the mass production of fraud and deception.

Since data can be stored, traded and used long after the transaction, future data use is likely to grow with data processing technology such as AI. More importantly, future data use is *obscure to both* sides of the transaction when the buyer decides whether to give away personal data in a focal transaction. The seller may be reluctant to restrict data use to a particular purpose, a particular data processing method or a particular time horizon, in light of future data technology. Even if it does not plan to use any data technology itself, it can always sell the data to those that will use it. These data markets motivate the seller to collect as much information as consumers are willing to give.

Sophisticated consumers may anticipate the uncertainty and hesitate to give away personal data. However, in many situations, identity and payment information are crucial (or made crucial) to complete the focal transaction, leaving even the most sophisticated consumers to trade off between immediate gains from the focal transaction and potential loss from future data use. One may argue that future data use is simply a new attribute of the product traded in the focal transaction; as long as the attribute is clearly conveyed between buyer and seller (say via a well-written privacy policy), sellers in a competitive market will respect buyer preference for limited data use. Unfortunately, this attribute is not currently well defined at the time of the focal transaction, and it can evolve over time in ways that depend on the seller's data policy but are completely out of the buyer's view, control, ability to predict or ability to value. This ongoing information asymmetry, if not addressed, could lead to a lemon's market (with respect to future data use).

Incomplete information about future data use is not the only problem lurking in the interaction between AI and consumer privacy. There are at least two other problems related to the uncertainty about future data use and value: one is externality, and the other is commitment.

To be clear, future data use can be beneficial or detrimental to consumers, thus rational consumers may prefer to share personal data to some extent (Varian 1997). However, benefits from future data use – e.g., better consumer classification, better demand prediction or better product design – can usually be internalized by the collector of the information via internal data use or through the sale of data to third parties. In contrast, damages from future misuse – e.g., identity theft, blackmail or fraud – often accrue not to the collector but to the consumer. Because it is often hard to trace back consumer harm to a particular data collector, these damages may not be internalized by either the data collector, or by consumers in their choices about how to interact with the collector. This is partly because the victim consumer may have shared the same information with hundreds of sellers, and she has no control how each piece of information may get into the wrong hands. The asymmetry between accruable benefits and non-accountable damages amounts to negative externality from sellers to buyers. If there is no way to track back to the origin, sellers have an incentive to over-collect buyer information.[1][2]

This difficulty in tracing damages back to actions by the data collector, together with uncertainty about future use and ongoing information asymmetry about collector practices, also triggers a commitment problem. Assuming consumers care about data use, every seller has an incentive to boast about having the most consumer-friendly data policy in the focal transaction, but will also retain the option to renege after data collection. There might be some room to enforce a declared data policy specific

[1] The argument of negative externality has been discussed in multiple papers, including Swire and Litan (1998) and Odlyzko (2003). See Acquisti et al. (2016) for a more comprehensive summary.

[2] There could be positive externality from one player to another. For example, a data set that tracks an infectious disease nationwide can generate enormous public health benefits for everyone. But if each data collector accesses only part of the data and there is no way for him to benefit from the final product based on nationwide data, he may have an incentive to under-collect and under-share the data. Here I focus on negative externality, in order to highlight the risk of over-collecting and over-sharing.

promises, if the seller's actual practice is revealed to the public and found to contradict its promise. However, it is often difficult to discover the real data practice. It is even more difficult to rectify consumer damage from a misrepresented data policy, as a court often requires a "body on the ground" – i.e., evidence of a harmful outcome – as well as some confidence that there is a causal link between that outcome and the data collector's practices.[3]

Information asymmetry, externality and commitment concerns can all be exacerbated by AI. More specifically, by potentially increasing the scope and value of consumer data use, AI can increase the expected benefits and costs of big data. But since the benefits are more internalized to the owner of the data and AI than consumer risks, AI could encourage intrusive use of data despite higher risks to consumers. For the same reason, new benefits enabled by AI – say cost savings or better sales – could entice a firm to (secretly) abandon its promise in privacy or data security.

In short, big data introduces three "new" problems for consumer privacy: (1) sellers initially have more information about future data use than buyers after the focal transaction; (2) sellers need not fully internalize potential harms to consumers because of the inability to trace harm back to a data collector; and (3) sellers may promise consumer-friendly data policy at the time of data collection but renege afterwards, as it is difficult to detect and penalize it ex post.[4] All three encourage irresponsible data collection, data storage and data use.

All three problems could be aggravated by AI and other data technologies. Later in the essay, I will describe a few AI-powered techniques that aim to *alleviate* the risk to consumer privacy and data security. Hence, the net impact of AI on privacy needs to take both sides into account.

2. Ongoing risk in consumer privacy and data security

[3] The Court's emphasis on tangible harm is best illustrated in an ongoing battle between the Federal Trade Commission (FTC) and LabMD. LabMD is a medical testing laboratory that collects sensitive personal and medical information from consumers. The FTC alleged that LabMD violated the FTC Act by failing to employ reasonable and appropriate measures to prevent unauthorized access to consumers' personal information. In November 2015, the Administrative Judge of the FTC dismissed the FTC complaint, arguing that complaint counsel failed to prove that LabMD's data security conduct caused or was likely to cause substantial injury to consumers (https://www.ftc.gov/news-events/press-releases/2015/11/administrative-law-judge-dismisses-ftc-data-security-complaint). This decision was reversed in July 2016, by an Opinion and Final Order from the FTC commissioners (https://www.ftc.gov/news-events/press-releases/2016/07/commission-finds-labmd-liable-unfair-data-security-practices). In November 2016, the 11th U.S. Circuit Court of Appeals granted LabMD's request to temporarily stop enforcing the FTC order (while the appeals court considers the case), on the grounds that mere emotional harm and actions causing only a low likelihood of consumer harm may not meet the legal definition of unfair practice, even when the exposed data is highly sensitive. The court opinion can be found at http://f.datasrvr.com/fr1/016/73315/2016_1111.pdf. What type of consumer harm is needed for a data security practice to be unfair and illegal remains an open question.

[4] Jin and Stivers (2017) elaborate on the three information problems in more details, but they do not associate them with AI or other data technology.

The risk associated with privacy and data security is real. Fundamentally data driven, the risk can be directly or indirectly related to AI and other data technologies. For example, since AI enhances the expected value of data, firms are encouraged to collect, store and accumulate data, regardless of whether they will use AI themselves. The ever-growing big data storehouses become a prime target to hackers and scammers.

2.1 Data at risk

According to the Privacy Rights Clearinghouse, 7,859 data breaches have been made public since 2005, exposing billions of records with personal identifiable information (PII) to potential abuse.[5] A closer look at the data is even more alarming: not only do we observe mega breaches that affect millions at once, but also the information lost in a single breach spreads to all kinds of PII. When Target lost 40 million records in December 2013, hackers got mostly debit and credit card numbers. But the recent Equifax breach (September 2017) affected 145 million people, with social security number, whole credit history, and even driver license and transaction dispute data stolen from the same database. More concerning is the fact that data breaches occur disproportionally to organizations that accumulate massive PII data, including retailers, information aggregators, financial institutions, and non-profit organizations such as governments, schools, and hospitals.

Causes of data breaches have evolved as well. A decade ago, most data losses were driven by human errors such as unshredded records left in the trash, lost laptops without encrypted data or data inadvertently uploaded to the open web. Recent breaches are often the result of targeted hacking and ransomware attack. If we view a malicious hacker as a thief sneaking in to steal, a ransomware attacker is a kidnapper who takes control of your data system and demands ransom immediately. For instance, the ransomware attack in May 2017 has infected computers in 99 countries (including the U.S.), bringing down transportation, banking, nuclear and hospital systems in many places.[6]

Thomas et al. (2017) follow the dark web from March 2016 to March 2017, passively monitoring forums that trade credential leaks exposed via data breaches, phishing kits that deceive users into submitting their credentials to fake login pages, and off-the-shelf keyloggers that harvest passwords from infected machines. They identify large number of potential victims, including 788,000 of off-the-shelf keyloggers, 12.4 million of phishing kits, and 1.9 billion usernames and passwords exposed via data breaches. After matching these exposed credentials to Google's internal database, they find that 7 to 25 percent of exposed passwords match a victim's Google account. More alarmingly, they observe "a remarkable lack of external pressure on bad actors, with phishing kit playbooks and keylogger capabilities remaining largely unchanged since the mid-2000s."

2.2 Consumers at risk

The most concrete harm that could arise from a data breach is identity theft. According to

[5] https://www.privacyrights.org/data-breaches, accessed on December 18, 2017.
[6] http://www.bbc.com/news/technology-39901382, accessed on October 20, 2017.

the Bureau of Justice Statistics (BJS), identity theft affects 17.6 million (7 percent) of all U.S. residents age 16 and older (Harrell 2014). Consistently, identity theft is one of the biggest consumer-complaint categories – first in 2014, second in 2015 and third in 2016 (FTC 2014, 2015, 2016). In 2016, identity theft accounted for 13 percent of consumer complaints, trailing behind debt collection (28 percent) and imposter scam (13 percent), all of which could feed on lost personal data (FTC 2016).

Of course, not all identity thefts are driven by inadequate privacy protection or insufficient data security. Scammers practiced their creative art long before big data and AI existed. However, loss from identity theft is likely a function of data misuse. As reported by BJS (Harrell 2014), 86 percent of identity theft victims experienced fraudulent use of existing account information and 64 percent reported a direct financial loss from the identity theft incident. Among those who reported direct financial loss, victims of personal information fraud lost an average of $7,761 (with a median of $2,000) and victims of existing bank fraud lost an average of $780 (with a median of $200).[7]

Researchers have attempted to draw a statistical link between data misuse and consumer harm. Romanosky, Acquisti and Telang (2011) explore differences among state data breach notification laws and find that adoption of data breach disclosure laws reduces identity theft caused by data breaches by an average 6.1 percent. Romanosky, Hoffman and Telang (2014) further examine federal data breach lawsuits from 2000 to 2010. They show that the odds of a firm being sued are 3.5 times greater when individuals suffer financial harm but 6 times lower when the firm provides free credit monitoring. Telang and Somanchi (2017) look at a more indirect consequence of data misuse. Using detailed transaction data from a US bank, they find that consumers are three percentage points more likely to leave the bank if they have experienced an unauthorized fraudulent transaction within six months. While the unauthorized transaction could be a result of previous data breaches, it is difficult to attribute the fraud to a particular data breach. In other words, the bank and the consumer may both suffer from a data breach, but the breached firm has virtually zero shares in this suffering.

Tax fraud offers another peek into the harm of data misuse. Through GAO (2015), the U.S. Internal Revenue Service (IRS) reported a point estimate of attempted identity theft refund fraud (as of 2013). Although the IRS was able to prevent or recover $24.2 billion in fraudulent refunds, it paid out $5.8 billion in tax refunds that were later flagged as identity theft frauds. In May 2015, the IRS disclosed a data breach where 100,000 taxpayer accounts were compromised through its Get Transcript application. This breach exposes sensitive information such as taxpayers' prior-year tax filings. More important, it is compromised not because hackers broke a digital backdoor of the IRS, but because hackers were able to clear a multi-step authentication process that required prior personal knowledge of the taxpayer's social security number, date of birth, tax filing status and street address.[8] In other words, hackers got in the front door of the IRS, using information

[7] Direct financial loss is not necessarily equal to the actual out-of-pocket loss to identity theft victims, as some financial loss may be reimbursed.

[8] https://www.irs.gov/newsroom/irs-statement-on-the-get-transcript-application, accessed on October 19, 2017.

they already had or could readily guess. Such information is likely from previous data breaches or data available on the black market. This suggests that data breaches could have a ripple effect: a small vulnerability in one database could undermine data security in a completely unrelated organization.

In some situations, data in the wrong hands could cause damage much bigger than fraudulent charges. For instance, the breach of AshleyMadison.com was said to be linked to multiple suicides.[9] The ransomware attack in May 2017 was reported to have shut down work in 16 UK hospitals[10], crippled medical devices[11], and delayed at least one surgery in a U.S. hospital.[12] As more medical devices get connected to the Internet, compromised data security could generate disruption in surgeries and life support. It is not difficult to imagine similar risks in connected cars and the "internet of things."

One may argue that the ongoing wave of data breaches is more driven by data availability than by data processing technology. This could be true at the moment, but recent trends suggest that criminals are getting sophisticated and are ready to exploit data technology.

For instance, robocalls – the practice of using a computerized auto-dialer to deliver a pre-recorded message to many telephones at once – has become prevalent because of relatively standard advances in information technology. But improved methods of pattern recognition and delivery appear to have increased the efficacy, and thus prevalence, of these calls. For example, by pretending the call is from a local number that looks familiar to the receiver, it tricks the receiver into listening to unwanted telemarketing. Similarly, phishing emails have long strived to target people vulnerable to financial and other frauds. Because the phishing attempt can be much more effective if it appears to come from a familiar email address and contains personal information that is supposedly only known to family and friends, effective phishing attempts have been limited by the labor needed to customize each email. This danger can be easily magnified when scammers mass produce PII-customized phishing emails with individualized targeting, appeals and mass delivery.

Ironically, the same data technology that giant tech firms use for legitimate business can be converted into a tool for data misuse. AI is no exception. On September 6, 2017, Facebook admitted that it received approximately $100,000 in ad revenue from roughly 3,000 ads connected to 470 inauthentic accounts and Pages that are affiliated with each other and likely operated out of Russia.[13] Such information was estimated to reach as

---

[9] http://www.dailymail.co.uk/news/article-3208907/The-Ashley-Madison-suicide-Texas-police-chief-takes-life-just-days-email-leaked-cheating-website-hack.html, http://money.cnn.com/2015/08/24/technology/suicides-ashley-madison/index.html, accessed on October 26, 2017.

[10] https://www.theverge.com/2017/5/12/15630354/nhs-hospitals-ransomware-hack-wannacry-bitcoin, accessed on October 20, 2017.

[11] https://www.forbes.com/sites/thomasbrewster/2017/05/17/wannacry-ransomware-hit-real-medical-devices/#7666463e425c, accessed on October 20, 2017.

[12] https://www.recode.net/2017/6/27/15881666/global-eu-cyber-attack-us-hackers-nsa-hospitals, accessed on October 20, 2017.

[13] https://newsroom.fb.com/news/2017/09/information-operations-update/, accessed on October 19, 2017.

many as 126 million U.S. users.[14] Similar discoveries followed from Twitter and Google. The ongoing investigation suggests that these Russian-backed accounts chose their content strategically so that the algorithms embedded in the platforms – including search rank, ad targeting and post recommendation – helped to broadcast the message to specific demographics.[15]

It is not going to be long before the same algorithms get exploited for stalking, blackmail and other shady use. According to Vines et al. (2017), one can spend as low as $1,000 to track someone's location with mobile ads. This is achieved by exploiting the ad tracking and ad targeting algorithms widely used in mobile platforms and mobile apps. We do not know whether this trick has been used in the real world, but it sends two chilling messages. First, personal data is not only available to giant consumer-facing companies that can use AI for mass, individualized *but impersonal*, marketing but is also within the reach of small, non-market parties who can exploit that data for *personalized* targeting of the consumer. Arguably, the latter is more dangerous to a targeted individual, as small non-market parties face less reputation constraint, they are invisible to consumers, and they may be interested in causing more harm than simply getting a consumer to purchase an unwanted product. Second, these bad actors may be able to take advantage of the key algorithms that are designed to reap the benefits of AI for legitimate purposes. As these algorithms are further developed, they could also empower data misuse.

Even if we can keep all data tightly secured and limit AI to its intended use, there is no guarantee that the intended use is harm free to consumers. Predictive algorithms often assume there is a hidden truth to learn, which could be the consumer's gender, income, location, sexual orientation, political preference or willingness to pay. However, sometimes the to-be-learned "truth" evolves and is subject to external influence. In that sense, the algorithm may intend to *discover* the truth but end up *defining* the truth. This could be harmful, as algorithm developers may use the algorithms to serve their own interest, and their interests – say earning profits, seeking political power, or leading cultural change – could conflict with the interest of consumers.

The danger of misleading algorithms is already seen in the controversy about how Russia-sponsored posts got disseminated in social medias during the 2016 U.S. presidential election. In the Congressional hearings held on October 31 and November 1, 2017, lawmakers expressed the concern that the business model of Facebook, Twitter and Google, which depends on advertising revenue from a large user base, may hamper their willingness to identify or restrict misinformation from problematic users.[16] Because social media users are more likely to consume information that platform algorithms push

[14] https://www.nytimes.com/2017/10/30/technology/facebook-google-russia.html, accessed on December 18, 2017.
[15] https://www.nytimes.com/2017/09/07/us/politics/russia-facebook-twitter-election.html, accessed on October 19, 2017. http://money.cnn.com/2017/09/28/media/blacktivist-russia-facebook-twitter/index.html, accessed on October 19, 2017.
[16] The full video and transcript of these hearings are available at c-span.org (https://www.c-span.org/video/?436454-1/facebook-google-twitter-executives-testify-russia-election-ads, and https://www.c-span.org/video/?436360-1/facebook-google-twitter-executives-testify-russias-influence-2016-election&live).

to them, they may end up consuming information that hurt them in the future.[17]

The same conflict of interest has sparked concerns in price discrimination. This argument is that if AI enables a firm to predict a consumer's willingness to pay, it could use that information to squeeze out every penny in consumer surplus. This argument is plausible in theory, but needs to be evaluated with at least three considerations: first, if more than one firms can use AI to discover the same consumer willingness to pay, competition among them will ease the concern of perfect price discrimination; second, the economics literature has long demonstrated the ambiguous welfare effect of price discrimination. As long as price discrimination is imperfect (i.e., firms cannot charge every consumer's willingness to pay), some consumers may benefit from the practice (via lower price) while other consumers suffer. From a social planner's point of view, whether to encourage or punish AI-enabled price discrimination depends on the weights it assigns to different parts of society. Third, in the long run, AI may reduce the operational costs within the firm (e.g., via a more cost-effective inventory management system) and foster product innovations that better fit consumer demand. These changes could be beneficial to both the firm and its consumers.

A somewhat opposite concern is that AI and other predictive technology are not 100 percent accurate in their intended use. It may not introduce much inefficiency or wasteful effort if Netflix cannot precisely predict the next movie I want to watch, but it could be much more consequential if the U.S. National Security Agency (NSA) flags me as a future terrorist based on some AI algorithm. As Solove (2013) has argued, it is almost impossible for someone to prove that they will not be a terrorist *in the future.* But at the same time, they may be barred from air travel, have personal conversation with friends monitored, and be restricted from work, trade and leisure activities. If this AI algorithm applies to a large population, it could do a lot of harm even if the probability of error is close to zero.

To summarize, there is a real risk in privacy and data security. The magnitude of the risk, and its potential harm to consumers, will likely depend on AI and other data technologies.

3. How does the U.S. market cope with the risk in privacy and data security?

Before we jump into a regulatory conclusion, we must ask how the market copes with the risk in privacy and data security. Unfortunately, the short answer is that we do not know much. Below I describe what we know on the demand and supply sides, along with a summary of existing public policies in the U.S. Admittedly, the literature cited below is more about privacy and data security than about AI. This is not surprising, as AI has just started to find its way into e-commerce, social media, national security and the internet of things. However, given the ongoing risk and the potential interaction of AI and that risk, it is important to keep in mind the big picture.

[17] Note that a predicative algorithm is not necessarily more biased than human judgment. For example, Hoffman, Kahn and Li (2017) study job-testing technologies in 15 firms. They find that hires made by managers against test recommendations are worse on average. This suggests that managers often overrule test recommendations because they are biased or mistaken.

3.1 Consumer Attitude

On the demand side, consumer attitude is heterogeneous, evolving, and sometimes self-conflicting.

When surveyed, consumers often express serious concerns about privacy, although self-reported value of privacy covers a wild range (see summary in Acquisti et al. 2016). However in real transactions, many consumers are willing to give away personal data in exchange for a small discount, free services, or a small incentive such as a pizza (Athey, Catalini and Tucker 2017). This conflict, which some referred to as a "privacy paradox," suggests that we have yet to comprehend the link between consumer attitude and consumer behavior. Furthermore, researchers have found that privacy preference varies by age (Goldfarb and Tucker 2012), by time (Stutzman, Gross and Acquisti 2013), and by context (Acquisti, Brandimarte and Loewenstein 2015). Although old data are shown to add little value to search results (Chiou and Tucker 2014), biometric data such as fingerprint, facial profiles and genetic profiles can be much longer lasting (Miller and Tucker 2017). Hence, consumers may have a different preference on biometric data than on the data that gets obsolete fast. These heterogeneities make it even harder to paint a complete picture of consumer attitude and consumer behavior about privacy.

A similar puzzle exists for attitudes towards data security. A recent survey by the Pew Research Center suggests that many people are concerned about the safety and security of their personal data in light of numerous high-profile data breaches (Pew 2016). However, according to Ablon, Heaton, Lavery and Romanosky (2016), only 11 percent stopped dealing with the affected company and 77 percent were highly satisfied with the company's post-breach response.

It is hard to tell why consumers are willing to give away data in real transactions. One possibility is that consumers have a large or even hyperbolical discount for the future, which motivates them to value the immediate gains from the focal transaction more than the potential risk of data misuse in the distant future. Other behavioral factors can be at play as well. Small incentives, small navigation costs and irrelevant but privacy-reassuring information can all persuade people to relinquish personal data, according to a recent field experiment (Athey, Catalini and Tucker 2017).

It is also possible that news coverage – on data breaches and privacy problems – raises consumer concern about the overall risk, but they do not know how to evaluate the risk specific to a transaction. Despite heavy news coverage, people may have an illusion that hacking will not happen to them. This illusion could explain why John Kelly, the former Department of Homeland Security head and current White House chief of staff, used a compromised personal phone for months.[18]

The third explanation is that consumers are fully aware of the risk, but given the fact that their personal data has been shared with many firms and has likely already been breached

[18] https://www.wired.com/story/john-kelly-hacked-phone/, accessed on October 15, 2017.

somewhere, they believe the extra risk of sharing the data with one more organization is small. Survey evidence seems to lend some support to this conjecture. According to Pew (2016), few are confident that the records of their activities maintained by various companies and organizations will remain private and secure. A vast majority (91 percent) of adults agree that consumers have lost control of how PII is collected and used by companies, though most think personal control is important. Moreover, 86 percent of Internet users have taken steps to remove or mask their digital footprints, and many say they would like to do more or are unaware of tools they could use.[19]

Consumer anxiety may explain why identity theft protection service has become a $3 billion industry (according to IBISWorld[20]). However, a market review by GAO (2017) shows that identity theft services offer some benefits but generally do not prevent identity theft or address all of its variations. For instance, these services typically do not address medical identity theft or identity theft refund fraud. In fact, a number of identity theft service providers were caught making deceptive marketing claims,[21] casting doubt on whether such "insurance-like" services are the best way to safeguard consumers from identity theft.

3.2 Supply side actions

Statistics from the supply side are mixed, too.

Thales (2017a) conducted a global survey of 1,100 + senior security executives, including 100+ respondents in key regional markets in the U.S., U.K. Germany, Japan, Australia, Brazil and Mexico, and in key segments such as Federal Government, Retail, Finance and Healthcare. It finds that 68 percent of survey respondents have ever experienced a breach, while 26 percent experienced one last year. Both numbers rose from 2016 (61 percent and 22 percent).

For financial services in particular, Thales (2017b) finds that firms are aware of the cyber risk they face but tend to deploy new technology (e.g., cloud, big data, internet of things) *before* adopting security measures to protect them. Only 27 percent of US financial services organizations said to feel 'very' or 'extremely' vulnerable to data threats (the

[19] "The state of privacy in post-Snowden America" by the Pew Research Center, source: http://www.pewresearch.org/fact-tank/2016/09/21/the-state-of-privacy-in-america/.

[20] https://www.ibisworld.com/industry-trends/specialized-market-research-reports/technology/computer-services/identity-theft-protection-services.html, accessed on October 26, 2017.

[21] For example, in September 2012, Discover settled with the Consumer Financial Protection Bureau (CFPB) and the Federal Deposit Insurance Corporation (FDIC) with $200 million refund to consumers and $14 million penalty. CFPB and FDIC alleged that Discover engaged in misleading telemarketing on identity theft protection, credit score tracking, wallet protection, and payment protection (http://money.cnn.com/2012/09/24/pf/discover-penalty-telemarketing/index.html.) In December 2015, LifeLock agreed to pay $100 million to settle FTC contempt charges for order violation. The 2010 court order requires the company to secure consumers' personal information and prohibits the company from deceptive advertising in identity theft protection services (https://www.ftc.gov/news-events/press-releases/2015/12/lifelock-pay-100-million-consumers-settle-ftc-charges-it-violated).

global average is 30 percent), despite the fact that 42 percent of US financials had been breached in the past (the global average is 56 percent). Consistently, both U.S. and global financials rank data security at the bottom of their spending plans, citing institutional inertia and complexity as the main reasons. These numbers should be concerning, because the financial sector has the highest cost of cybercrime according to the latest report from Accenture (2017). To add a little comfort, Thales (2017b) also reports that security spending, which includes but is not limited to data security, continues to trend up: 78 percent of U.S. financials reported higher spending than last year, trailing only US Healthcare (81 percent) and ahead of the overall global average (73 percent).

Firms' willingness to invest in data security is partially driven by the cost they suffer directly from data breaches. A strand of literature has studied the stock market's response to data breach. While results differ across studies, the general finding is that the financial market response is small and temporary, if negative at all (Campbell et al. 2003; Cavusogru et al. 2004; Telang and Wattal 2007; Ko and Dorantes 2006). A couple of studies have provided an absolute estimate of the cost. According to Ponemon (2017), who surveyed 419 organizations in 13 countries and regions, the average consolidated total cost of a data breach is $3.62 million. Ponemon (2017) further finds that data breaches are most expensive in the U.S., with the average per capita cost of data breach as high as $225. In contrast, Romanosky (2016) examines a sample of 12,000 cyber events, including but not limited to data breaches. He finds that the cost of a typical cyber incident (to the affected firm) is less than $200k, roughly 0.4 percent of the firm's estimated annual revenues.

Thousands or millions, these estimates only reflect the direct cost of the cyber event to *the firm*, not all the consequential harm *to consumers*. For example, most breached firms offer one-year free credit monitoring service for affected consumers, but data misuse can occur after a year. Either way, consumers have to spend time, effort and money to deal with identity theft, reputation setback, fraud, blackmail or even unemployment as a result of a data breach. The lawsuit between FTC and Wyndham Hotel and Resort gives a concrete example. Wyndham was breached multiple times in 2008 and 2009, affecting more than 619,000 consumers. Before reaching a settlement, the FTC alleged that fraudulent charges attributable to the Wyndham breaches exceeded $10.6 million.[22] Although the final settlement involves no money, this case suggests that harm to consumers – via an increased risk of identity theft and the costs to mediate the risk – can be much more substantial than the direct loss suffered by the breached firm. Arguably, it is this difference that motivates firms to over-collect data or use lax data security, despite the real risk of data breach.

The good news is that, market forces do push firms to respect consumer demand for privacy and data security. For instance, Facebook profiles expand over time and therefore the same default privacy setting tends to reveal more personal information to larger

[22] https://www.washingtonpost.com/business/economy/2012/06/26/gJQATDUB5V_story.html?utm_term=.1ab4fedd7683, accessed October 19, 2017.

audiences.[23] In September 2014, Facebook adjusted its default setting of privacy from public posting to friend-only posting, which limits third party access to new users' Facebook posts. In the meantime, Facebook made it easier for existing users to update their privacy settings, block out ads, and edit their ad profiles.[24] We do not know the exact reason behind the change, but it could be related to a few things: for example, user willingness to share data on Facebook dropped significantly from 2005 to 2011(Stutzman, Gross and Acquisti 2013), academic research shows that it is very easy to identify strangers based on photos publicly posted on Facebook (Acquisti, Gross and Stutzman 2014) and it costs Facebook $20 million to settle a class action lawsuit regarding its "sponsored stories" (an advertising feature alleged to misappropriate user profile pictures and likenesses without user consent).[25]

Similarly, a privacy scare prompted Samsung to change its privacy policy. In February 2015, CNN quoted a paragraph of Samsung's privacy policy, which stated that words spoken in front of a Samsung Smart TV are captured and transmitted to a third party through use of voice recognition.[26] In response to the intense fear that smart TVs "spy" in a private living room, Samsung later changed its privacy policy.[27] Samsung also clarified that voice recognition can be disabled and it uses industry standard encryption to secure the data.

The privacy competition in the smartphone market is even more interesting. In 2015, Google launched Android Marshmallow in Android 6.0,[28] which prompts users to grant or deny individual permissions (e.g., access to the camera) to a mobile app when it is needed for the first time, rather than automatically grant apps all of their specified permissions at installation. It also allows users to change the permissions at any time. Similar features were made available earlier in Apple iOS 8.[29] Apple's commitment to privacy protection was also highlighted when Apple refused to unlock the iPhone from one of the shooters in the December 2015 terrorist attack in San Bernardino, California.

As a pioneer in biometric authentication, Apple recently announced Face ID in its next smartphone launch (iPhone X). Using infrared cameras, Face ID uniquely identifies a

---

[23] Matt McKeon gives a graphical account of how Facebook privacy evolves from 2005 to 2010, at http://mattmckeon.com/facebook-privacy/, accessed on October 24, 2017.
[24] http://60secondmarketer.com/blog/2014/09/21/facebook-tightens-privacy-controls-affect-marketing/, accessed October 24, 3017.
[25] https://www.wired.com/2013/08/judge-approves-20-million-facebook-sponsored-stories-settlement/, accessed October 24, 2017.
[26] According to CNN (http://money.cnn.com/2015/02/09/technology/security/samsung-smart-tv-privacy/index.html), Samsung's privacy policy said "Please be aware that if your spoken words include personal or other sensitive information, that information will be among the data captured and transmitted to a third party through your use of Voice Recognition." The article further points out that, Samsung SmartTV has a set of pre-programmed commands that it recognizes even if you opt out of voice recognition.
[27] https://www.cnet.com/news/samsung-changes-smarttv-privacy-policy-in-wake-of-spying-fears/, accessed on October 24, 2017.
[28] Android Mashmellow was first released as a beta on May 28, 2015, followed by the official release on October 5, 2015. Its new model of app permission was received positively: https://fpf.org/2015/06/23/android-m-and-privacy-giving-users-control-over-app-permissions/.
[29] https://fpf.org/2014/09/12/ios8privacy/, accessed on October 24, 2017.

user's face and utilizes that information to unlock the smartphone and authorize Apple Pay. Though it is meant to enhance convenience and security, Face ID has stirred a number of privacy concerns including exposing consumer privacy to Apple employees and allowing the police to forcefully unlock a phone using the owner's face. Whether this AI-powered technology will reduce or enhance privacy protection is an open question.

Note that market mechanisms can also work *against* consumer privacy and data security. Dina Florêncio and Cormac Herley (2010) examined the password policy of 75 websites and found that password strength is *weaker* for some of the largest, most attacked sites that should have greater incentives to protect their valuable database. Compared to security demand, it seems that competition is more likely to drive websites to adopt a weaker password requirement, as they need to compete for users, traffic and advertising. The sample size of this study is too small to represent the whole market, but the message is concerning: consumer demand in privacy and data security may compete with the same consumers' demand for convenience, usability, and other attributes (such as lower price). When these demands conflict with each other, firms may have a stronger incentive to accommodate the attributes that are more visible and easier to evaluate. Probably the same reason explains why only a small fraction of firms adopt multi-factor authentication[30], despite its ability to reduce data risk.

So far, we have considered AI as an external factor that potentially increases the risk of privacy violation and data breach. It is important to recognize that AI could also serve as *a tool to mitigate* the risk. Recently, AI has demonstrated super intelligence in games such as Go, even without the help of any human knowledge (Silver et al. 2017). Imagine what data risk would look like if the same AI power is used to grant data access to authorized personnel, to detect data attack when (or even before) it materializes, and to precisely predict whether a user-generated posting is authentic or fake. In fact, the technology frontier is moving this direction, though its net benefits remain to be seen.

Take differential privacy as an example. It was invented more than 10 years ago (Dwork et al. 2006) and claimed by Apple as a key feature to protect consumer identity in some of its data collection since 2016. The basic logic goes as follows: the data collecting firm adds random noise to an individual user's information before uploading it to the cloud. That way, the firm can still use the collected data for meaningful analysis without knowing each user's secret. The effectiveness of this technology depends on how much noise to add, a parameter under the control of the data-collecting firm.

To evaluate how Apple implements differential privacy in practice, Tang et al. (2017) reverse-engineered Apple's MacOS and iOS operating systems. They find that the daily privacy loss permitted by Apple's differential privacy algorithm exceeds values acceptable by the theoretical community (Hsu et al. 2014), and the overall privacy loss per device may be unbounded. Apple disputes the results and argues that its differential privacy feature is subject to user opt-in. Google is another user of differential privacy (in

[30] Multifactor authentication is a security measure that requires two or more independent credentials to verify the identity of the user. https://twofactorauth.org/ allows one to search whether a firm uses multi-factor authentication in various types of products or services.

its web browser Chrome). The "noise" parameter that Google uses – as estimated by Erlingsson, Pihur and Kolonova (2014) – seems to be more privacy-protective than what is claimed to be used in Apple, but still falls short of the most-acceptable range.[31] These debates cast doubt on the promise of differential privacy, especially on its real use relative to its theoretical potential.

Another promising technology is blockchain. In plain English, blockchain is an ever-growing list of records (blocks) that are linked with timestamp and transaction data. Secured by cryptography, blockchain is designed to be verifiable, permanent, and resistant to data modification. Its successful application in bitcoin suggests that similar technology could trace identities in data trade and data use, thus reducing the risk in privacy and data security (Catalini and Gans 2017). Ironically, a ransomware attacker in May 2017 demanded bitcoin instead of traditional money, probably for a similar security reason.

3.3 Policy landscape

Any market description is incomplete without a summary of the policy background. In the U.S., there is no overarching legislation on consumer privacy or data security. So far, the policy landscape is a patchwork of federal and local regulations.

Only a few federal laws are explicit on privacy protection and they all tend to be industry specific. For example, the Gramm-Leach-Bliley Act (GLBA) controls the ways that financial institutions deal with personal data; the Health Insurance Portability and Accountability Act of 1996 (HIPPA) provides data privacy and security provisions for medical records; and the Children's Online Privacy Protection Act of 1998 (COPPA) disciplines online services directed to children under the age of 13. In accordance, privacy is subject to federal regulation by sectors: the Department of Health & Human Resources (DHHS) enforces HIPPA in health care, the Federal Communication Commission (FCC) regulates telecommunication services, the federal reserve systems monitors the financial sector, the Security and Exchange Commission (SEC) focuses on public firms and financial exchanges, and the Department of Homeland Security (DHS) deals with terrorism and cybercrimes related to national security.

Two exceptions are worth mentioning. First, the Federal Trade Commission (FTC) can address privacy violations and inadequate data security as deceptive and unfair practice, following the 1914 FTC Act. This enforcement authority covers almost every industry and overlaps with many sector-specific regulators.

More specifically, FTC's privacy enforcement focuses on "notice and choice", which emphasizes how firms' actual data practice deviates from the privacy notice they disclose to the public. For industries not subject to GLBA, HIPPA or COPPA, there is no legislation that mandates privacy notice, but many firms provide it voluntarily and seek consumer consent before purchase or consumption. Some industries also adopt self-

[31] https://www.wired.com/story/apple-differential-privacy-shortcomings/, accessed on October 24, 2017.

regulatory programs to encourage certain privacy practices.[32] This background allows the FTC to obtain privacy notice of the targeted firm and enforce it under the FTC Act.

The FTC has published a number of guidelines on privacy,[33] but the best way to understand its enforcement is through cases. For example, the FTC alleged that Practice Fusion misled consumers by first soliciting reviews for their doctors and then publicly posting these reviews on the internet without adequate consumer notice. The case eventually settled in June 2016.[34] In another case against Vizio, FTC alleged Vizio captured second-by-second information about video displayed on its smart TV, appended specific demographic information to the viewing data, and sold this information to third parties for targeted ads and other purposes. According to the complaint, VIZIO touted its "Smart Interactivity" feature that "enables program offers and suggestions" but failed to inform consumers that the settings also enabled the collection of consumers' viewing data.[35] The case is joint with New Jersey Attorney General and settled for $2.2 million in February 2017. The third case is against Turn, a digital advertising company that tracks consumers in online browser and mobile devices, and uses that information to target digital advertisements. The FTC alleged that Turn used unique identifiers to track millions of Verizon consumers even after they choose to block or delete cookies from websites, which is inconsistent with Turn's privacy policy. Turn settled with FTC in December 2016.[36]

While privacy notice is something that consumers can access, read (whether they read them is another question) and consent to, most data security practices are not visible until someone exposes the data vulnerability (via data breach or white-hat discovery). Accordingly, FTC enforcement on data security focuses on whether a firm has adequate data security, not whether the firm has provided sufficient information to consumers. Following this logic, the FTC has settled with Ashley Madison, Uber, Wyndham Hotel and Resorts, Lenovo and TaxSlayer, but is engaged in litigation with LabMD and D-Link.[37]

The second exception relates to government access to personal data. Arguably, the U.S. Constitution, in particular the First and Fourth Amendments, has already covered individual rights in free speech and limited government ability to access and acquire

[32] For example, Digital Advertising Alliance (DAA), a non-profit organization led by advertising and marketing trade associations, establishes and enforces privacy practices for digital advertising.
[33] The most comprehensive FTC guideline is its 2012 privacy report (FTC 2012). A list of privacy-related press releases can be found at https://www.ftc.gov/news-events/media-resources/protecting-consumer-privacy/ftc-privacy-report.
[34] https://www.ftc.gov/news-events/press-releases/2016/06/electronic-health-records-company-settles-ftc-charges-it-deceived, accessed on October 24, 2017.
[35] https://www.ftc.gov/news-events/press-releases/2017/02/vizio-pay-22-million-ftc-state-new-jersey-settle-charges-it, accessed on October 25, 2017.
[36] https://www.ftc.gov/news-events/press-releases/2016/12/digital-advertising-company-settles-ftc-charges-it-deceptively, accessed on October 25, 2017.
[37] For a list of FTC cases in data security, see https://www.ftc.gov/enforcement/cases-proceedings/terms/249.

personal belongings. However, exactly how the Constitution applies to electronic data is subject to legal debate (Solove 2013).

Beyond the Constitution, a series of federal laws – the Electronic Communications Privacy Act of 1986 (ECPA), the Stored Communications Act (1986), the Pen Register Act (1986), and the 2001 USA Patriot Act – stipulate when and how the government can collect and process electronic information of individuals. But many of these laws were enacted in the wake of the Watergate scandal, long before the use of the Internet, email, search engines, and social medias. It is unclear how they apply to real cases. The legal ambiguity is highlighted in three events: first, as exposed by Edward Snowden, the NSA has secretly harvested tons of personal information for its global surveillance programs. The exposure generates an outcry for privacy and a hot debate in the balance between individual privacy and national security. Second, the U.S. Supreme Court has yet to hear the Microsoft email case, regarding whether the U.S. government has the right to access emails stored by Microsoft overseas.[38] Third, Apple refused to unlock the iPhone of one of the shooters in the 2015 San Bernardino terrorist attack. Since the FBI was able to unlock the phone before the court hearing, it remains unknown whether Apple has the legal obligation to help the FBI.[39]

At the local level, 48 of the 50 states have enacted data breach notification laws, bur no federal law has been passed on this topic.[40] According to the National Conference of State Legislatures, at least 17 states have also passed some law on privacy. These local laws tend to vary greatly in content, coverage and remedy.[41] From the research point of view, these variations are useful for studying the impact of data breach laws on identity theft (Romanosky, Acquisti and Telang 2011)[42] and data breach lawsuits (Romanosky,

[38] http://www.reuters.com/article/us-usa-court-microsoft/u-s-supreme-court-to-decide-major-microsoft-email-privacy-fight-idUSKBN1CL20U, accessed October 25, 2017.
[39] http://www.latimes.com/local/lanow/la-me-ln-fbi-drops-fight-to-force-apple-to-unlock-san-bernardino-terrorist-iphone-20160328-story.html, accessed October 25, 2017.
[40] There have been multiple efforts towards a federal data breach notification law. In 2012, Senator Jay Rockefeller advocated for a cybersecurity legislation that strengthens the requirement to report cybercrimes. In January 2014, the Senate Commerce, Science and Transportation Committee (led by Senator Rockefeller) introduced a bill to create a federal requirement for data breach notification (S. 1976 Data Security and Breach Notification Act of 2014). In his 2015 State of the Union Speech, President Obama proposed new legislation to create a national data breach standard with a 30-day notification requirement for data breach. A related bill was later introduced by the US House of Representatives (H.R. 1770L Data Security and Breach Notification Act of 2015). All of them failed. In the wake of the mega breaches in 2017, Congress has introduced Personal Data Notification and Protection Act of 2017 (H.R. 3806), the Data Protection Act of 2017 (H.R. 3904), the Market Data Protection Act of 2017 (H.R. 3973), Cyber Breach Notification Act (H.R. 3975), Data Broker Accountability and Transparency Act (S. 1815) and Data Security and Breach Notification Act (S. 2179). They are under committee review and likely consolidated.
[41] The National Conference of State Legislatures collects information on these state laws. For data breach laws, see http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx. For privacy laws, see http://www.ncsl.org/research/telecommunications-and-information-technology/state-laws-related-to-internet-privacy.aspx.
[42] Romanosky, Acquisti and Telang (2011) explore differences among state data breach notification laws and link them to a FTC database of identity theft from 2002 to 2009. They find that adoption of data breach disclosure laws reduces identity theft caused by data breaches by an average 6.1 percent.

Hoffman and Telang 2014), but they can be difficult to comply if a firm operates in multiple states. It is also difficult for consumers to form an expectation of privacy protection, especially if they transact with both in-state and out-of-state firms.

In short, the U.S. system is piecemeal and multi-layered, in contrast to the European Union's attempt to unify data protection via its General Data Protection Regulation (effective in 2018).[43] Which approach is better for society is subject to an ongoing debate.

4. Future challenges

To summarize, there are pressing issues in consumer privacy and data security, many of which are likely to be reshaped by AI and other data technologies.

A number of big questions arise: shall we continue to let the market evolve under the current laws, or shall we be more aggressive in government regulation? How do firms choose data technology and data policy if consumers demand both convenience and privacy? How to balance AI-powered innovations against the extra risk that the same technology brings to privacy and data security? If action is needed from policy makers, shall we let local governments use trial and error, or shall we push for federal legislations nationwide? Shall we wait for new legislations to address standing loopholes, or shall we rely on the court system to clarify existing laws case by case? These questions deserve attention from researchers in many disciplines, including economics, computer science, information science, statistics, marketing, and law.

In my opinion, the leading concern is that firms are not fully accountable for the risk they bring to consumer privacy and data security.[44] To restore full accountability, one needs to overcome three obstacles, namely (1) the difficulty to observe firms' actual action in data collection, data storage and data use; (2) the difficulty to quantify the consequence of data practice, especially before low-probability adverse events realize themselves; and (3) the difficulty to draw a causal link between a firm's data practice and its consequence.

These difficulties exist, not only because of technical limits, but also because of misaligned incentives. Even if blockchain can track every piece of data and AI can predict the likelihood of every adverse event, whether to develop and adopt such technology is up to firms. In the current setting, firms may still have incentives to hide real data practice from the public, to obfuscate information disclosed to consumers, or to blame other random factors for consumer harm.

Changes must be made to instill more transparency into the progression from data practice to harmful outcomes, and to translate outcomes (realized or probabilistic) into incentives that directly affect firms' choice of data practice. These changes should not aim to slow down data technology or to break up big firms just because they are big and on the verge of an AI breakthrough. Rather, the incentive correction should aim to help

[43] An overview of GDPR is available at http://www.eugdpr.org/.
[44] The same problem applies to non-profit organizations and governments.

consumer-friendly data practice stand out from lemons, which in turn fosters innovations that respect consumer demand for privacy and data security.

There might be multiple ways to address misaligned incentives, including new legislation, industry self-regulation, court ruling, and consumer protection. Below I comment on the challenges of a few of them.

First, it is tempting to follow the steps in safety regulation. After all, the information problems we encounter in privacy and data security – as highlighted in Section 1 – are similar to those in food, drug, air, car or nuclear safety. In those areas, the consequence of inadequate quality control is random and noisy, just as identity thefts and refund frauds are. In addition, firm input and process choices – like ingredients and plant maintenance – are often unobservable to final consumers. A common solution is direct regulation on the firm's action: for example, restaurants must keep food at a certain temperature; nuclear plants must pass periodical inspections, etc. These regulations are based on the assumption that we know what actions are good and what actions are bad. Unfortunately, this assumption is not easy to come by in data practice. With fast evolving technology, are we sure that politicians in Washington, DC are the best ones to judge whether multi-factor authentication is better than a 20-character password? How do we ensure that the regulation is updated with every round of technological advance?

The second approach relies on firm disclosure and consumer choice. "Notice and choice" is already the backbone of FTC enforcement (in privacy), and data breach notification laws follow a similar principle. For this approach to be effective, we assume consumers can make the best choice for themselves as long as they have adequate information at hand. This assumption is unlikely to hold in privacy and data security, because most consumers do not read privacy notices (McDonald and Cranor 2008), many data-intensive firms may not have a consumer interface, and it could be difficult for consumers to choose as they do not have the ability to evaluate different data practices and do not know what choices are available to mitigate the potential harm. Furthermore, firms' data practice may change frequently in light of technological advance, thus delivering updated notices to consumers may be infeasible and overwhelming.

The third approach is industry self-regulation. Firms know more about data technology and data practice, and therefore are better positioned to identify best practices. However, can we trust firms to impose and enforce regulations on themselves? History suggests that industry self-regulation may not occur without the threat of government regulation (Fung et al. 2007). This suggests that efforts pushing for government action may be complementary rather than substitutable to industry attempts to self-regulate. Another challenge is technical: many organizations are trying to develop a rating system on data practice, but it is challenging to find comprehensive and updated information firm by firm. This is not surprising, given the information asymmetry between firms and consumers. Solving this problem is crucial for any rating system to work.

The fourth approach is defining and enforcing privacy and data use as "rights." Law scholars have long considered privacy as a right to be left alone, and debated whether

privacy rights and property rights should be treated separately (Warren and Brandeis 1890). As summarized in Acquisti et al. (2016), when economists consider privacy and data use as rights, they tend to associate them with property rights. In practice, the EU has followed the "human rights" approach, which curtails transfer and contracting rights that are often assumed under a "property rights" approach. The EU recognized individual rights of data access, data processing, data rectification and data erasure in the new legislation (GDPR, to be effective in 2018). The impact of GDPR remains to be seen, but two challenges are worth mentioning: first, for many data-intensive products (say self-driving cars), data do not exist until the user interacts with the product, often under third-party support (say GPS service and car insurance). Should the data belong to the user, the producer, or third parties? Second, even if property rights over data can be clearly defined, it does not imply perfect compliance. Music piracy is a good example. Both challenges could deter data-driven innovations, if the innovator has to obtain the rights to use data from multiple parties beforehand.

Apparently, no approach is challenge free. Given the enormous impact that AI and big data may have on the economy, it is important to get the market environment right. This environment should respect consumer demand for privacy and data security, encourage responsible data practices and foster consumer-friendly innovations.

References:

Ablon, Lilian; Paul Heaton; Diana Lavery and Sasha Romanosky (2016): *Consumer Attitudes Toward Data Breach Notifications and Loss of Personal Information*. Santa Monica, CA: RAND Corporation, 2016.

Accenture (2017): *2017 Insights on the Security Investments That Make a Difference*. Available at https://www.accenture.com/t20170926T072837Z__w__/us-en/_acnmedia/PDF-61/Accenture-2017-CostCyberCrimeStudy.pdf.

Acquisti, Alessandro; Curtis Taylor and Liad Wagman (2016) "The Economics of Privacy", *Journal of Economic Literature*, 54(2): 442-92.

Acquisti, Alessandro, Allan Friedman, and Rahul Telang (2006) "Is There a Cost to Privacy Breaches? An Event Study." Presented at *the Twenty-Seventh International Conference on Information Systems*, Milwaukee 2006 and *Workshop on the Economics of Information Security 2006*.

Acquisti, Alessandro; Gross, Ralph; and Stutzman, Fred (2014) "Face Recognition and Privacy in the Age of Augmented Reality," *Journal of Privacy and Confidentiality*: Vol. 6: Iss. 2, Article 1. Available at: http://repository.cmu.edu/jpc/vol6/iss2/1

Acquisti, Alessandro; Laura Brandimarte and George Loewenstein (2015) "Privacy and human behavior in the age of information" *Science*, January 30, 2015; 347(6221): 509-14.

Athey, Susan and Catalini, Christian and Tucker, Catherine E., "The Digital Privacy Paradox: Small Money, Small Costs, Small Talk" (September 27, 2017). *MIT Sloan Research Paper* No. 5196-17; *Stanford University Graduate School of Business Research Paper* No. 17-14. Available at SSRN: https://ssrn.com/abstract=2916489 or http://dx.doi.org/10.2139/ssrn.2916489

Campbell, Katherine; Lawrence A. Gordon; Martin P. Loeb and Lei Zhou (2003): "The Economic Cost of Publicly Announced Information Security Breaches: Empirical Evidence from the Stock Market." *Journal of Computer Security* 11 (3): 431–48.

Catalini, Christian and Gans, Joshua S., "Some Simple Economics of the Blockchain (September 21, 2017)" *Rotman School of Management Working Paper* No. 2874598; *MIT Sloan Research Paper* No. 5191-16. Available at SSRN: https://ssrn.com/abstract=2874598 or http://dx.doi.org/10.2139/ssrn.2874598

Cavusoglu, Huseyin; Birendra Mishra and Srinivasan Raghunathan (2004) "The Effect of Internet Security Breach Announcements on Market Value: Capital Market Reactions for Breached Firms and Internet Security Developers." *International Journal of Electronic Commerce* 9 (1): 69–104.

Chiou, Lesley and Catherine E. Tucker (2014). Search engines and data retention: Implications for privacy and antitrust.

Dwork, Cynthia; Frank McSherry; Kobbi Nissim; and Adam Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference* (TCC). 265–284.

Erlingsson, Úlfar; Vasyl Pihur and Aleksandra Korolova. 2014. RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security* (CCS). 1054–1067.

FTC (2012) *Protect Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policy Makers*, available at https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

FTC (2014): *Consumer Sentinel Network Data Book from January – December 2014*, accessed at https://www.ftc.gov/system/files/documents/reports/consumer-sentinel-network-data-book-january-december-2014/sentinel-cy2014-1.pdf.

FTC (2015): *Consumer Sentinel Network Data Book from January – December 2015*, accessed at https://www.ftc.gov/system/files/documents/reports/consumer-sentinel-network-data-book-january-december-2015/160229csn-2015databook.pdf.

FTC (2016): *Consumer Sentinel Network Data Book from January – December 2016*,

accessed at https://www.ftc.gov/system/files/documents/reports/consumer-sentinel-network-data-book-january-december-2016/csn_cy-2016_data_book.pdf.

Florêncio, Dina and Cormac Herley (2010) "Where Do Security Policies Come From?" *Symposium on Usable Privacy and Security (SOUPS)* 2010, July 14–16, 2010, Redmond, WA USA.

Fung, Archon; Mary Graham, and David Weil *(2007) Full Disclosure: The Perils and Promise of Transparency*, Cambridge University Press.

Goldfarb, Avi and Catherine E. Tucker (2012). Shifts in privacy concerns. *American Economic Review: Papers and Proceedings* 102(3), 349–53.

Government Accountability Office (GAO): *Identity Theft Services: Services Offer Some Benefits but Are Limited in Preventing Fraud*, GAO-17-254, March 2017, available at http://www.gao.gov/assets/690/683842.pdf.

Government Accountability Office (GAO): *Identity Theft and Tax Fraud: Enhanced Authentication Could Combat Refund Fraud, but IRS Lacks an Estimate of Costs, Benefits and Risks*, GAO-15-119, January 2015. Available at https://www.gao.gov/products/GAO-15-119.

Harrell, Erika (2014): *Victims of Identity Theft, 2014*, Bureau of Justice Statistics, available at https://www.bjs.gov/content/pub/pdf/vit14.pdf.

Hsu, Justin; Marco Gaboardi; Andreas Haeberlen; Sanjeev Khanna; Arjun Narayan; Benjamin C Pierce and Aaron Roth. 2014. Differential privacy: An economic method for choosing epsilon. In *27th IEEE Computer Security Foundations Symposium (CSF).* 398–410.

Jin, Ginger Zhe and Andrew Stivers, "Protecting Consumers in Privacy and Data Security: A Perspective of Information Economics" (May 22, 2017). *Working paper*, available at SSRN: https://ssrn.com/abstract=3006172.

Ko, Myung, and Carlos Dorantes (2006) "The Impact of Information Security Breaches on Financial Performance of the Breached Firms: An Empirical Investigation." *Journal of Information Technology Management* 17 (2): 13–22.

McDonald, Aleecia, and Lorrie Faith Cranor. 2008. "The Cost of Reading Privacy Policies." *I/S: A Journal of Law and Policy for the Information Society* 4 (3): 540–65.

Miller, Amalia and Catherine E. Tucker (2017). Privacy protection, personalized medicine and genetic testing. Forthcoming *Management Science*.

Odlyzko, Andrew. 2003. *Privacy, Economics, and Price Discrimination on the Internet." In Economics of Information Security*, edited by L. Jean Camp and

Stephen Lewis, 187–212. Norwell, MA: Kluwer Academic Publishers.

Pew Research Center (2016): *The state of privacy in post-Snowden America*, available at http://www.pewresearch.org/fact-tank/2016/09/21/the-state-of-privacy-in-america/.

Ponemon (2017): *2017 Ponemon Cost of Data Breach Study*, available at https://www.ibm.com/security/data-breach/index.html.

Posner, Richard A. 1981. "The Economics of Privacy." *American Economic Review* 71 (2): 405–09.

Romanosky, Sasha; David Hoffman and Alessandro Acquisti (2014), "Empirical Analysis of Data Breach Litigation," *Journal of Empirical Legal Studies*, 11(1):74-104, 2014.

Romanosky, Sasha; Alessandro Acquisti and Rahul Telang (2011), "Do Data Breach Disclosure Laws Reduce Identity Theft?" *Journal of Policy Analysis and Management*, 30(2):256-286, 2011.

Romanosky, Sasha (2016): "Examining the costs and causes of cyber incidents" *Journal of Cybersecurity*, Volume 2, Issue 2, 1 December 2016, Pages 121–135, https://doi.org/10.1093/cybsec/tyw001

Silver, David; Julian Schrittwieser; Karen Simonyan; Ioannis Antonoglou; Aja Huang; Arthur Guez; Thomas Hubert; Lucas Baker; Matthew Lai; Adrian Bolton; Yutian Chen; Timothy Lillicrap; Fan Hui; Laurent Sifre; George van den Driessche; Thore Graepel and Demis Hassabis "Mastering the game of Go without human knowledge", *Nature*, 550, pp 354–359 (19 October 2017).

Solove, Daniel (2013), *Nothing to Hide: The False Tradeoff between Privacy and Security*, Yale University Press.

Stigler, George J. 1980. "An Introduction to Privacy in Economics and Politics." *Journal of Legal Studies* 9(4): 623–44.

Swire, Peter P., and Robert E. Litan (1998): *None of Your Business: World Data Flows, Electronic Commerce, and the European Privacy Directive*. Washington, DC: Brookings Institution Press.

Tang, Jun; Aleksandra Korolova; Xiaolong Bai; Xueqiang Wang and Xiaofeng Wang (2017) "Privacy Loss in Apple's Implementation of Differential Privacy on MacOS 10.12", available at https://arxiv.org/pdf/1709.02753.pdf.

Telang, Rahul, and Sunil Wattal (2007) "An Empirical Analysis of the Impact of Software Vulnerability Announcements on firm Stock Price." *IEEE Transactions on Software Engineering* 33 (8): 544–57.

Telang, Rahul and Sriram Romanchi (2017) “Security, Fraudulent transactions and Customer Loyalty: A Field Study”, *working paper*.

Thales (2017a): *2017 Thales Data Threat Report: Trends in Encryption and Data Security (Global Edition)*, available at https://dtr.thalesesecurity.com/.

Thales (2017b): *2017 Thales Data Threat Report: Trends in Encryption and Data Security (Financial Services Edition).* Available at https://dtr-fin.thalesesecurity.com/.

Thomas, Kurt; Frank Li; Ali Zand; Jacob Barrett; Juri Ranieri; Luca Invernizzi; Yarik Markov; Oxana Comanescu; Vijay Eranti; Angelika Moscicki; Daniel Margolis; Vern Paxson; and Elie Bursztein (2017): “Data Breaches, Phishing, or Malware? Understanding the Risks of Stolen Credentials”, Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pages 1421-1434. available at https://acmccs.github.io/papers/p1421-thomasAembCC.pdf.

Varian, Hal R. 1997. “Economics Aspects of Personal Privacy.” In *Privacy and Self-Regulation in the Information Age*. Washington, DC: US Department of Commerce, National Telecommunications and Information Administration.

Vines, Paul; Franziska Roesner and Tadayoshi Kohno (2017): “Exploring ADINT: Using Ad Targeting for Surveillance on a Budget—or—How Alice Can Buy Ads to Track Bob”, *the 16th ACM Workshop on Privacy in the Electronic Society* (WPES 2017).

Warren, Samuel and Louis Brandeis, “The Right to Privacy,” *Harvard Law Review*, vol. 4 (1890), p. 191.