LUOHAN
ACADEMY

UNDERSTANDING BIG DATA

DATA CALCULUS IN THE DIGITAL ERA

2021

The authors include:

Patrick Bolton
Columbia University

Long Chen
Luohan Academy

Bengt Holmström
Massachusetts Institute of Technology

Eric Maskin
Harvard University

Sir Christopher Pissarides
London School of Economics

Michael Spence
Stanford University

Tao Sun
International Monetary Fund

Tianshu Sun
University of Southern California

Wei Xiong
Princeton University

Liyan Yang
University of Toronto

UNDERSTANDING BIG DATA: DATA CALCULUS IN THE DIGITAL ERA

February 5, 2021

List of Authors

Luohan Community

Patrick Bolton

Bengt Holmström

Eric Maskin

Sir Christopher Pissarides

Michael Spence

Tao Sun

Tianshu Sun

Wei Xiong

Liyan Yang

In-house

Long Chen

Yadong Huang

Yong Li

Xuan Luo

Yingju Ma

Shumiao Ouyang

Feng Zhu

[*] This version is updated on February 5, 2021. Click here for the newest version.
We thank, for their thoughtful insights and comments, Nageeb Ali, Benjamin Bai, Xiao Cheng, James Dempsey, Nico van Eijk, Isabelle Falque-Pierrotin, Fuping Gao, Clarisse Girot, Jane Horvath, Rohit Lamba, Haiying Li, Weiqiu Long, Wei Lv, Ariane Mole, Zhengjun Nie, Pietro Ortoleva, Thomas Sargent, Jean Tirole, Laura Veldkamp, Neng Wang, Stephen Kai-Yi Wong, Hanhua Zhou, and all participants in the 2019 Privacy and Data Governance Workshop in Hangzhou, the 2nd Luohan Academy Annual Conference, the Privacy Workshop (Princeton-Luohan Academy), and various internal workshops. We also thank Thomas Walton and Michael Whinihan for their edits of the draft as well as helpful advice.

Mission Statement of Luohan Academy

Digital technology is fundamentally changing our global economy with the potential to advance human welfare in many ways. It not only reduces costs and market frictions, but also enables the development of new services and processes. Consumers may benefit in numerous ways, from lower costs to improved services, and from greater social connectivity to better health outcomes. Entrepreneurs and firms (especially small firms) may benefit from having low-friction access to marketplaces, cheaper computing and back-office services, as well as earlier and cheaper, more efficient sources of financing. Governments may be able to lower the costs of administering their welfare systems, better anticipate and serve the needs of their citizens, and deliver services more efficiently and inclusively. Moreover, by building a valuable network for producers and consumers, digital platforms can generate new opportunities for resource and risk allocation, and provide a stage for efficient and resilient routines, processes restructuring, as well as market relationships.

At the same time, our society is not yet well-prepared for this unprecedented structural transformation brought by big data, machine learning, artificial intelligence, robotics, and other digital technologies. Thus, it is imperative that we study and manage the coming digital revolution to benefit society and protect individuals as consumers, workers, and citizens, both domestically and internationally.

Citizens in all countries are confronted with numerous questions about the optimal and balanced use of these new technologies.How can societies harness the power of technology to promote growth, enhance societal welfare and at the same time preserve individual rights and social inclusion? What is the future of work and leisure? How must education change to address the needs of the changing nature of work, the rise of digital assistance, the rapid change of technology and new methods enabled by the digital revolution? How to avoid a “digital illiteracy” that results in “digital knowledge gap” across citizens? How can we ensure that the new social environments will be fair and inclusive? Which appropriate regulation and competition policies foster competition and technological progress without slowing innovation? What are the contours of a privacy policy that will allow legitimate use of data to create socially beneficial and inclusive services, while protecting citizens against abuses by unauthorized agents and institutions? How can digital technology contribute to a greener planet?

Social scientists in general, including economists, must therefore collaborate to help societies adapt smoothly and fairly to the digital revolution.Two important objectives of the academic community are first, to understand business models and market structures that enable growth and progress, and second, to identify the impact of digitization on individual and social welfare. So far the rapidly increasing scale of digitization has not been followed by a corresponding increase in theoretically grounded empirical research on the rationales, consequences, and policies of digitization. A well-organized research community could greatly facilitate and speed up such research efforts.

This is an opportune time to bring the best research minds in the world together with first hand practical insights into the digital economy to advance the research frontiers of digitization and shape constructive consensus for the public good.The Luohan Academy thus has a two-fold mission. The first is to understand how digital technology can help achieve the common good. The second is to help build a broad research community for systematic and in-depth research leading to new paradigms for solving first-order problems in the digital society. In this endeavor, the academy will abide by the spirit of open science, operating independently under the principles of integrity, inclusion and diversity.

It is an exciting new beginning. We sincerely invite you to join.

Foreword

Information processing has been defined as “the change (processing) of information in any manner detectable by an observer, a process that describes everything that happens in the universe, from the falling of a rock to the printing of a text file from a digital computer system.” Human society has long realized that information processing and sharing are essential to the pursuit of physical, social, and economic well-being. Knowing and sharing information about one’s surroundings is essential to success in the physical world. Knowing and sharing information about one’s “neighbors” is essential to success in the social world. “Knowing-your-customers” (KYC) in order to serve them well is the supreme norm for success in the business world. Sharing that information – the Yellow Pages practice that makes certain personal information, such as name, telephone number, and address public -- has become a tradition in modern human relations.

The pervasive use of digitized information has reached a new height that we call the era of "big data." While this has led to unprecedented societal cooperation, it has also intensified three major concerns: ***How can we properly protect personal privacy in the age of big data? How do we understand and manage the ownership and distribution of benefits and risks arising from the use of data? Will the use of big data lead to "winner-take-all" markets that undermine competition to the detriment of consumers and society?*** These are the subjects of this report.

We focus on analyzing concrete evidence about "big data" to draw conclusions on its impact. As Nobel Laureate Ronald Coase (1994) suggested, it is important to step away from pure "blackboard economics" that tends to only live in [a theoretician's] mind: "what we need is more empirical work ... An inspired theoretician might do as well without such empirical work, but ... the inspiration is most likely to come through the stimulus provided by the patterns, puzzles, and anomalies revealed by the systematic gathering of data, particularly ***when the prime need is to break our existing habits of thought."***

This viewpoint is particularly relevant because, unlike many production inputs, data has the properties of non-rivalry and non-separability. Unless an evidence-based, integrated and multi-stakeholder approach is adopted, users can be unintentionally hurt in the name of protection. We don’t want to "dismember the goose that laid the golden egg."

Chapter 1.
Introduction and Summary

> ***"We are in the era of big data. With a smartphone now in nearly every pocket, a computer in nearly every household, and an ever-increasing number of Internet-connected devices in the marketplace, the amount of consumer data flowing throughout the economy continues to increase rapidly."***
>
> ***-- United States Federal Trade Commission (FTC), January, 2016***

1.1 An age defined by digital data

Over millennia, human beings have learned to gather, organize, store, and share vast amounts of complex information. But there is more knowledge in the world than can ever be gathered and distributed collectively and correctly, let alone processed by all those who stand to benefit from the exercise. To complicate matters, self-motivated individuals or organizations may, intentionally or unintentionally, provide incorrect or incomplete information in the pursuit of their own interests. Nobel Laureate Friedrich Hayek went so far as to suggest that "the economic problem of society is ... a problem of the utilization of knowledge not given to anyone in its totality." This is because "the knowledge of the circumstances of which we must make use never exists in concentrated or integrated form, but solely as disbursed bits of incomplete, and frequently contradictory knowledge which all the separate individuals possess"(Hayek, 1945). Hayek makes information processing the key challenge in the pursuit of human economic well-being.

Many economists since then have expanded and elaborated on this idea. Another Nobel Laureate, Douglas North, said that "the fundamental theoretical problem underlying the question of cooperation is the manner by which individuals attain knowledge of each other's preferences and likely behavior" (North, 1990). The more a producer understands her customers, the better she can satisfy their demand. In the United States, basic personal data, including name, address, and phone number are shared publicly through the Yellow Pages to enhance connections among people. In areas such as commerce, medicine and finance, "knowing your customer" (KYC) is a prerequisite for a mutually beneficial relationship. This involves sharing and storing private and sensitive information.

Information is "asymmetric." People engaged with each other possess different types of information. And they either do not or cannot convincingly restore symmetry by freely exchanging their information. This asymmetry causes departures from economic efficiency. Asymmetry can be a major obstacle to efficient transactions (Spence, 1973, 1974; Grossman and Stiglitz, 1980). When the problem is severe enough, a market can disappear (Akerlof, 1970). In the labor market, lack of information about workers' abilities and firms' needs is a cause of departures from efficient resource allocation. And these are manifested in unemployment and low productivity (Phelps, 1970; Pissarides, 2000).

Such considerations led Ronald Coase (1994) to conclude that a lot of what we think of as economic activity is "designed to accomplish what high transaction costs would otherwise prevent or to reduce transaction costs so that individuals can negotiate freely and we can take advantage of that diffused-knowledge of which Friedrich Hayek has told us."

Many efforts have been made in promoting information collection and diffusion. Mechanisms have been designed to reduce transaction costs. And incentives have been used to encourage cooperation when economic actors are faced with insufficient or asymmetric information (see, for example, Hart, 1988; Hart and Moore, 1988; and Holmström, 1979, 1982).

In the 21st century, in the middle of a digital revolution, there is now more easily accessible information than ever before. It has become much easier to use the information for our benefit. But this has also enhanced the risks to information privacy and security. The digital revolution has transformed the role of information in society and the economy, elevating both its importance as a growth engine and concerns with how it might be used by others to our detriment. It owes its origins to the brilliant ideas of Claude Shannon and Alan Turing, who, in the 1940s, encoded data into its "digital atoms," known as bits. Combined with the new semiconductor technology used to store and process large volumes of data, these ideas led to an explosion of data. By the 1970s, the use of the word "data" surpassed that of "information."

We think it is important to start with a common understanding of what we mean by "data" and "big data." ***To be clear, data is not the same as information.*** Data are a collection of observations of some things. "Big data" relates to the assembling, storage, and processing of "(little) data." Much of "data science" is concerned with "data compression" – taking a large data set and reducing it to a much smaller set that contains most of the "information" that is then rendered easier to store and interpret.

The new millennium has been marked by the increasing use of "big data." Conceptualizing the term, "big data," means comprehending the three "V's": volume, variety, and velocity. ***"Volume"*** refers to the amount of data that can be processed and analyzed in order to understand and predict consumer behavior. ***"Variety"*** refers to the breadth of data that can be efficiently analyzed in order to satisfy the demands of sellers and buyers operating on digital markets. ***"Velocity"*** is the speed with which the data can be collected, processed, analyzed, and put to use.

The pervasive use of digitized data has intensified three major concerns:***How can we properly protect personal privacy in the age of big data? How do we understand the ownership and distribution of benefits and risks arising from the use of data? Will the use of big data lead to a situation of "winner-take-all" markets that undermine competition to the detriment of consumers and society?***

These are the issues that we address in this report. We are at a critical crossroads, a situation in which data has never been more important, yet where there is little consensus on how the use of data should be governed in order to address the equally important risk of abuse. To realize the full benefits of digital data, we need to better understand the nature of the data involved, how data are actually used, and what tradeoffs occur in controlling access to data. And we need to separate established facts from speculations and fears, founded and unfounded. A better understanding of the ***nature of data is what we call the "data calculus,"*** the subject of our inquiry.

We focus on analyzing concrete evidence using "big data" to draw conclusions on the impact of "big data." As Nobel Laureate Coase (1994) suggested, it is important to step away from pure "blackboard economics" that tends to only live in [a theoretician's] mind: "what we need is more empirical work ... An inspired theoretician might do as well without such empirical work, but ... the inspiration is most likely to come through the stimulus provided by the patterns, puzzles, and anomalies revealed by the systematic gathering of data, particularly ***when the prime need is to break our existing habits of thought."***

1.2 Understanding the nature of data

1.2.1 Understanding the privacy paradox

We start Chapter 2 with a simple, commonly accepted definition of information or data privacy: "[M]aintaining control over one's personal information," what U.S. Supreme Court Justice Louis Brandeis, called, "the right to be left alone" (Pavlou, 2011). From this vantage point, we then examine how people make decisions when they must provide certain personal information in order to enjoy the benefits of mobile services. The well-known "privacy paradox" says that, while most people say they care about privacy, they often appear to be too ready to share their information for free, or for a small amount of compensation. This seeming contradiction between people's presumed concerns over a fundamental right to privacy and their revealed behavior is a pattern that holds true across many different cultures.

There exist several resolutions to the privacy paradox. Many argue that either the person involved is not informed about the possible harm coming from privacy violations, or that she simply has to tolerate certain privacy violations in order to use popular software applications (Apps) because of a lack of other options (Chen and Michael, 2012). But today many such options are emerging, later in this report. An alternative, perhaps more plausible interpretation, holds that when facing real-life decisions, it is people's actual behavior and not what they say about their privacy concerns that reveals how they truly perceive the tradeoff between privacy and data benefits.

The key is to understand users' behavior when they do have options as they respond in the marketplace. As further discussed in the body of this report, we have conducted a large-scale empirical analysis using Alipay data. Alipay currently has over one billion active users in China, with millions of pop-up Apps called mini programs embedded on the Alipay platform. These mini programs belong to businesses that range from small startups to large corporations. Access to these Apps requires a user's permission to release certain personal data. Users can choose to opt-out later by withdrawing permission. The Apps differ in their degrees of sensitivity of the requested data and in the necessity of the requested data for their services. Combining these differences with user characteristics and choices, ***we provide the largest big-data study thus far on privacy-related decisions.***

What happens when users have options to choose whether they are willing to share certain personal information in order to obtain services provided by various mini programs? As in other countries, the average Chinese citizen is concerned about privacy. But when given options, she overwhelmingly chooses to share certain personal information in order to reap benefits from new digital services. More than 75% of users opted into mini programs when requested, with a fairly low and declining cancelation rate (0.12% per month) later, suggesting that they in general do not regret their choices. They knowingly trade off information sensitivity for higher quality of services. And they respond to privacy-related incidents. The accumulation of digital experience over time helps them to choose more carefully, and it ultimately helps them embrace more digital services in the long run. These patterns hold regardless of gender, age, or education.

1.2.2 The value of data: an economic interpretation of the three “Vs”

The fact that the majority of users are willing to share personal information with providers of services reveals their preference to enjoy the benefits of digital services. This naturally raises the question we address in Chapter 3 where ***We examine the value of online data sharing in three domains: connectivity, decision making, and trust.***

As we demonstrated in ***Digital Technology and Inclusive Growth*** (Luohan Academy, 2019), data sharing enhances connectivity. Because digital data are so easy to produce and share, there has been an unprecedented level of inclusive connectivity that has reshaped the marketplace and how people coordinate production and exchange. A case in point is the enormous increase in the scope, depth, and breadth of trading, made possible by online market exchanges. Offline trade has long been universally described by the "gravity model," in which the majority of customers in a local market came from within a ten-kilometer radius. The picture on the Taobao App for online trade stands in stark contrast. Each month more than 720 million active users shop there. They are served by more than ten million startups and companies. Half of these entrepreneurs are women. In terms of products, over three billion listed commodities and services can be ordered online. In terms of distance, except for fresh food, the average shopping distance between a buyer and seller is close to 1,000 kilometers, two orders of magnitude larger than the historical norm. The traditional shackles of trade within the gravity model have been broken.

This raises the question: What would happen if no personal data were used? Specifically, what would happen to the online market if its users were denied access to information flows recommended to them based on characteristics inferred from their personal data? We answer this question by conducting a large-scale randomized field experiment, switching off the personal data of 620,000 users in Taobao's recommendation algorithms. We find that the lack of such information has a dramatic impact on buyers and sellers alike. Without such personal data, no tailoring of services is possible, and platform recommendations blindly converge to the brands ranked in the top 1% of the goods and services exchanged, a pattern regularly observed in the pre-digital age. Without these data, transactions fall by a whopping 86%, with a disproportionally large adverse impact on small and micro-enterprises. The main point that emerges from this experiment is that matching users' data to products radically reduces search costs over a very large number of available products. If this information "spigot" were turned off and the flow of personal data shut down, buyers could rely only on traditional sources for information about prospective purchases: brand name, reputation, and general characteristics. Our experiment shows that, still today, these other sources are far from complete, so the size of the market is vastly reduced. This is consistent with the search literature, which shows that even small search or matching costs can make a big difference to the breadth and depth of both product and labor markets (Stigler, 1961, 1962; Diamond, 1971; Pissarides, 2009).

Second, ***data sharing enhances decision making.*** Large amounts and varieties of data, combined with connectivity, are enabling countless customers and producers to make smarter decisions, leading to more rapid and more beneficial product innovations, new and more innovative sales and services, and new and more innovative business models—new methods of industrial organization—that were simply not possible before (Luohan Academy, 2019) (See also the discussion of Schumpeterian competition in Chapter 6). This is particularly relevant for micro, small, and medium-sized companies (MSMEs) and individuals that, until now, have had little or no access to much-needed product and customer information. One example is Alibaba's Business Advisor, a service available to all online store owners with various information-analytical tools, such as sellers' historical performance, market trends, and the nature of their competitors. New subscribers, most of whom are MSMEs, typically enjoy a significant jump in sales growth within the first week of subscription, and the difference in performance between the subscribed and unsubscribed groups increases steadily over the next ten weeks. In addition, "big data" also helps MSMEs to grow by equipping them with a wide array of sophisticated analytical tools that used to be available only to large corporations.

Small businesses also benefit from the availability of finance without physical collateral, solving a heretofore insurmountable barrier to inclusive finance. This opens opportunities for tens of millions of entrepreneurs - a central finding of Luohan Academy's report. Big data has enabled the emergence of new large-scale microloans that were simply not possible before. Since 2011, MyBank has served more than 20 million MSMEs without collateral. They use the now famous "310" model: (less than) three minutes to apply for a loan, one second to obtain it, and zero personnel to complete the transaction. Information has become the new collateral in the digital age (Holmström, 2018). Information about the borrower is enough to assure a lender that it is a risk worth taking.

Third, ***data sharing builds trust.*** Trust in products and in other participants is essential for operating the new type of marketplace in which hundreds of millions of people make deals with each other, almost as if they were doing so face to face in the same local market (Tadelis, 2003). With online data sharing, customers increase their abilities to rate commodities and producers. That makes producers want to build their reputations as encoded in such ratings. All participants produce and benefit from such data exchanges – in sharp contrast to offline "lemon markets" where buyers lack the information that sellers have about the goods and services they want (Akerlof, 1970). Just as data benefits shoppers, it allows higher-quality repeat sellers to distinguish themselves from low-quality, "fly-by-night" sellers, shoring up their "brand," and benefitting from stronger sales over time.

Big data's larger Volume and greater Variety are fundamentally transforming online interactions and cooperation. They do so by changing how consumers and producers connect, by increasing trust between buyer and seller, and by facilitating better and more rapid decision-making. Crucially, this is happening in real-time at unheard of Velocity. Unlike finished commodities, data's value can only be materialized when data flows are being used. The three V's show us where the value of data comes from: large amounts of data and rich variety of data flows in real-time drive economic activities, confirming Hayek's insight regarding the benefits of decentralized decision making in open and competitive markets.

1.2.3 Risks of data sharing and possible solutions

As valuable as data sharing is, it is not without risks, the subject of Chapter 4. The more valuable data is, the greater the prospect for and urgency in protecting privacy and data safety. Every stage of the "life cycle" of big data is vulnerable to data breaches and privacy risks, from initial collection through compilation, analysis, and end use. While it is widely accepted that individuals should have the right to know and consent to data collection, in practice, it is a challenging task to protect individuals from excessive or even unauthorized disclosures. The number of data breaches and exposed records worldwide reached 1.6 billion in 2017, causing great consumer privacy concerns as well as huge economic losses. Incidents such as the Facebook – Cambridge Analytica data scandal have drawn widespread attention.

Effective privacy protection couples privacy engineering with privacy-enhancing technologies (PETs). Privacy engineering refers to a "privacy-by-design" approach that creates and implements software applications so that they can encourage privacy protection by service providers. It also designs user interfaces as to enhance users' understanding of privacy clauses and their ability to control sensitive information. For example, at the data collection stage, user authorization must be obtained, and technology companies must verify whether the requested data are indeed necessary. In the compilation and storage stages, before any data can be used, they can be desensitized, with sensitive information filtered out. They are encrypted so that they cannot be used in the event of data leakages. Only desensitized and encrypted data should be analyzed and used for understanding consumer preferences and predicting their behaviors, and this should be under close privacy risk management. Sustainable and practical use should strike a balance between the need for privacy protection and for minimizing user authorization – a balance that must reflect the wishes of the users. Finally, users must have deletion rights.

PETs are techniques and tools that provide privacy protection from untrustworthy and potentially harmful data controllers. One example is multiple-party calculation (MPC) technology, which allows analysts to gain insights from data without revealing or tracing back to the original data. Blockchain technology can also mitigate privacy risk by applying locks and keys to particular types of personal data. The goal is to provide big data analysis so that service providers know consumer characteristics to serve them well, "without knowing who they are." The downside of PETs is that they are difficult and costly to put into place, imposing additional challenges to startups and MSMEs in dealing with their customers.

A key takeaway is that, as time goes on, we can expect that data privacy and security issues will be mitigated to a large extent through the proper design of mechanisms and increasingly competent technologies. Some experts believe ***data sharing and the control of privacy and security risk do not have to constitute an insurmountable tradeoff.*** If so, then if technology is competent enough, the most serious and challenging privacy risks could become a thing of the past, if not entirely eliminated.

1.2.4 A simple framework for data analysis

In Chapter 5, we propose a simple framework for evaluating the tradeoffs between sharing and control in the digital age – a “data calculus.” Any framework must be based on an understanding of how users decide to share personal information in order to enjoy the benefits of online services. Secondly, it must factor in privacy and security risks, and how they can be mitigated through mechanism design and technology. Third, it must also reflect how the value of data is materialized as it is used in social and economic activities. Our data calculus framework incorporates the two general properties of any set of data, as viewed from three perspectives, and operating under one foundational principle. While the three V's are what make data special in the digital age, they do not change the essence of how data sharing works.

There are two general properties of data: Non-rivalry and non-separability. Unlike a physical commodity, once produced, data can be used an unlimited number of times. Also, irrespective of who uses the personal data, that very use can potentially have an impact on the data subjects.

We analyze these two properties from three different perspectives – that of data producers, data subjects, and use cases. Data producers are parties (in a business context) that collect, process, or control data. Data subjects refer to individuals (for personal data) or objects the data is about (for non-personal use). Data use cases refer to the economic or social activities where data is used.

In this framework, the interests of data producers and subjects are intertwined. It takes both data producers and data subjects to generate the data, and the use of data can impact both data producers and subjects. Data are costly to collect, process, and analyze for the producer. Yet the non-rivalry feature of data implies that there can be an unlimited number of owners without using up the data. For example, the information a speaker conveys in a speech is separately "produced" by the work of every participant’s eyes and ears; the information can then be passed on to people outside the meeting without consuming the information provided by the speaker.

From the data subjects' perspective, the use of data means their rights must be protected. The non-rivalry feature of data, and the intertwined interests of producers and subjects thus suggest three conclusions: ***First, it is inappropriate to think of data as a commodity with sole ownership. The ownership rights of all parties possessing data, so long as it is obtained through mutually-agreed methods, should be protected. Second, privacy protection is more about protecting the privacy rights of data subjects while their data is being used rather than granting them exclusive ownership. And third, there should be voluntary agreements such that both data producers and subjects can benefit from producing and using data.***

From the third, use case, perspective, we see that data cannot be treated as a commodity with a fixed value. In practice, the value of data is tied to the added value generated by economic and social interactions, as also predicted by the seminal work of Blackwell (1953). Accordingly, ***a central objective of data governance is to promote data flow, while protecting the rights of data subjects.*** This is the foundational principle that applies to all data, not just digitized data.

1.3 Data governance issues

1.3.1 The evolution of data governance

In Chapter 6 we show how our analytical framework can help the reader to better understand the evolution of data-privacy regulation and privacy governance. Modern privacy protection builds on Fair Information Practices (FIPs) originated in the early 1970s at the United States Department of Health, Education, and Welfare (HEW). They were based on five principles: (1) Notice/Awareness; (2) Choice/Consent; (3) Access/Participation; (4) Integrity/Security; and (5) Enforcement/Redress (United States Federal Trade Commission, 1998).

These principles became the basis of later guidelines and laws on privacy and personal data governance, including the Federal Trade Commission's effort to "encourage and facilitate effective self-regulation as the preferred approach to protecting consumer privacy online (FTC, 1998)" and the EU Data Protection Principles Directive, the General Data Protection Regulation (GDPR), and the California Consumer Privacy Act (CCPA).

A key goal of the FIP-based governance of personal data has been that it should not seek to lock up data or restrict ownership. Instead, consistent with the data calculus, the FIP principles aim to promote secure and privacy-protected data flow. For example, the OECD aims to "harmonize privacy legislation and, while upholding such human rights....at the same time prevent interruptions in…flows of data," in effect allowing consumers to benefit from the provision of their private information.

While no legislation restricts sole ownership, some rules exist that differ by restrictions on data use, reflecting the different views on how data should be governed. Though well intended, as pointed out by Nobel Laureate Elinor Ostrom, tight governance policies could have negative consequences for patents, intellectual property, licensing, pricing, as well as the ***"preservation of the digital economy."*** She argues that a better policy would rely more on multi-stakeholder policies that take advantage of the knowledge possessed by regulated companies (Layton, 2019, emphasis added). Similar findings and conclusions have been made by Goldfarb and Tucker (2011), Martin and Murphy (2016), and many other authors. "[R]egulation imposes its own costs, and digital privacy laws can restrict valuable information flows, increase privacy and security risks, erect barriers to market entry, increase uncertainty for entrepreneurs, and incite rent-seeking" (Layton, 2019).

1.3.2 Data and competition

Given that data-driven businesses play an increasingly important role in the economy, it is important to understand how their market practices affect competition. The broad objective is "to make sure that consumers benefit from the forces of competition" (Shapiro, 2018). Determining whether competition is so distorted that consumers may be harmed, requires an in-depth understanding of the structure of the industry, of business practices, and an evaluation of effective market performance.

We begin with a reflection on the positive influence of trade on competition and on the competitiveness of a nation's business enterprises. The advent of online trading has expanded China's trading radius by two orders of magnitude – from 10 to 1,000 kilometers on average. In ***The Wealth of Nations,*** published in 1776, Adam Smith points out that such expansions contribute to a "breaking down" of monopoly power, what he called a "great enemy to good management":

> "Good roads, canals, and navigable rivers, by diminishing the expense of carriage, put the remote parts of the country more nearly upon a level with those in the neighborhood of the town. They are upon that account the greatest of all improvements. They encourage the cultivation of the remote, which must always be the most extensive circle of the country. ***They are advantageous to the town, by breaking down the monopoly of the country in its neighborhood.*** They are advantageous even to that part of the country. ***Though they introduce some rival commodities into the old market, they open many new markets to its produce. Monopoly, besides, is a great enemy to good management, which can never be universally established but in consequence of that free and universal competition which forces everybody to have recourse to it for the sake of self-defence"*** (Smith, 1776; spelling in original English, emphasis added).

Just as "good roads, canals, and navigable rivers" extended the reach and broke down monopolies and increased the competitiveness of village traders in the time of Adam Smith, so has the emergence of digital networks in the 21st century "broken down" local monopolies and, as we will see, contributed to a stronger and more competitive business climate, albeit with much greater speed and intensity

Yet the possibility of harmful, anti-competitive business practices in the tech industries is garnering increasing attention and debate around the world. It is beyond the scope of this report to provide answers to all the many questions that have been raised, but we briefly discuss three key data-related questions for which we have some preliminary evidence.

Do businesses use big data to discriminate against consumers? Digital technology has transformed the relationships between producers and consumers. Sellers can obtain a heretofore unimaginable volume and diversity of information about their customers. One effect of the accumulation of this more granular data has been a shift from seeking to make as much money as possible on every individual product and service, to providing a user-centered bundled service, and to nurturing customer loyalty. Inclusion, namely having a large number and diversity of consumers, has become a business priority. For example, Ichihashi (2020) has shown that digital platforms prefer to disclose information about customer characteristics to sellers on the platform rather than to engage in what economists call "price discrimination" – selling an identical good or service at different prices to different groups of customers, extracting the highest price from those most willing to pay.

So far, while incidents always exist, there is little evidence around the world suggesting that big data accumulated by digital platforms has been mainly used to charge higher prices to consumers. In order to be successful, price discrimination requires the ability to separate out different groups one from the other, and the Internet has made that much more difficult than before. It enables buyers to search over many more product sellers across much greater distances than ever before, forcing much greater competition than ever before.

Price discrimination is not the only concern. "Companies can use big data to exclude low-income and underserved communities from credit and employment opportunities" and this is a concern of consumer protection agencies in the U.S., such as the Federal Trade Commission, which prosecutes such unfair trading practices (FTC, 2016). The increasingly inclusive nature of digital finance should mitigate that concern, at least in Kenya, China, and many parts of the world where such dramatic progress in inclusivity has taken place.

Has big data distorted competition and led to "winner-take-all" markets? Many of the concerns about competition and big data center on barriers that might arise from sizeable network externality effects (direct or indirect) and from huge economies of scale, giving rise to "winner-take-all" outcomes. Although concentration in the tech industry has been increasing in some sectors[1] in the U.S., the reality in China tells a different story. There, data-driven markets are characterized by low entry barriers and fierce competition. Incumbent firms are constantly vigilant about potential competitors.

While online consumption has exceeded 25% of all retail sales, the competition is intense and market shares have become less and less concentrated. In the space of just four years, even as Alibaba grows in an increasingly competitive way, its share of China's e-commerce sales has fallen from 78% in 2015 to 56% in 2019 due to new market entrants and to the growth of existing competitors. Pinduoduo, as[2] a new startup, attracted more than four hundred million users and grew sales by a factor of more than a hundred within three years. Baidu, China's long-dominant search engine and one-time leader in big data and artificial intelligence, had a larger market cap than Tencent and Alibaba in 2010. But now it is trailing far behind these two companies. ByteDance, the parent company of TikTok, came from nowhere and only took several years to take over Baidu in advertising revenue. JD.com, with its 17% of China's e-commerce sales and which enjoys the financial backing of American giants Google and Walmart, recently became "the platform with the largest market share of all channels in the home appliance market." A similar, intensely competitive trend is also happening in the mobile payment market. It is becomingly increasingly difficult to conclude that big data is leading to "winner-take-all" markets.

[1] https://www.emarketer.com/content/alibaba-jd-com-lead-in-china-but-a-few-others-are-making-dents-too and https://www.emarketer.com/content/retail-and-ecommerce-sales-in-china-2018

[2] https://en.wikipedia.org/wiki/List_of_largest_Internet_companies, https://supchina.com/2020/08/07/the-biggest-ecommerce-companies-in-china-a-brief-guide/, and https://equalocean.com/briefing/20200728230002802

Other factors beyond the use of big data diminish the risk of dominance and monopoly power. First, "data" is only a part of industrial organization in the digital economy. While digital technology has amplified the role of data in the business model, competition is still largely shaped by the underlying business model through which producers must compete for their customers. And unlike in physical space, users of digital services can have "multi-homing." That is, users can choose multiple different providers for similar services and spread their data around the Internet. Second, the marginal benefit of using data is decreasing with additional data. New data is generated constantly through economic activities, reducing the relevance of past data. As argued by Lambrecht and Tucker (2017), for a resource to provide a company with a competitive advantage, it must be inimitable, rare, valuable, and sustainable. Data is often none of these.

Is big data a barrier to innovation? Can an incumbent's control of big data be used to discourage innovation by new entrants? Although such concerns have been raised, we can think of several reasons why big data is a powerful driver for innovative products and production processes. First, the three V's of data have become an important driver of new methods of production and their requisite business models. They have amplified the ability of companies to connect with and understand their customers, to make smarter decisions, and to experiment with innovation. As documented in Digital Technology and Inclusive Growth (Luohan Academy, 2019) in almost every industry in which digital technology has played a transformational role, be it media, social media, e-commerce, finance, digital video, taxi-hailing, sharing bicycles, etc., the common denominator has been the development of new business models by new entrants. These new business models are powered by digital technology and data - "creative destruction" that displaces established models and companies that try to hang onto them. Almost everywhere around the world, the new models are often built on radical innovations by new entrants with little initial capital or other resources (except for boundless imagination and aspiration). Innovation is in the DNA of all tech start-ups.

Second, the three V's of data have led to an unprecedented level and scope of collaboration, which accelerates the roll-out of market-wide innovations. In particular, platforms have become important promoters and disseminators of innovation, both within and across affiliated institutions. It is in the interest of platforms, no matter their past successes, to foster innovation and competition. That is how their footprints expand and how they maintain interest in the platform. Perhaps the most striking aspect of innovation is the explosive growth of new niche brands on platforms.

The rest of this report is structured as follows: Chapter 2 presents the empirical evidence on consumer behavior in privacy-related decisions. Chapter 3 highlights the evidence on the value of data. Chapter 4 describes practices that can mitigate privacy concerns. Chapter 5 offers an in-depth, integrated framework for understanding data. Chapter 6 provides a closer look at some of the issues involved in data governance, focusing on the search for a middle ground between governmental and industry self-regulation, as well as an examination of the most recent theoretical and empirical evidence regarding the impact of "big data" on competition, innovation, and price discrimination. Chapter 7 offers some concluding thoughts and observations.

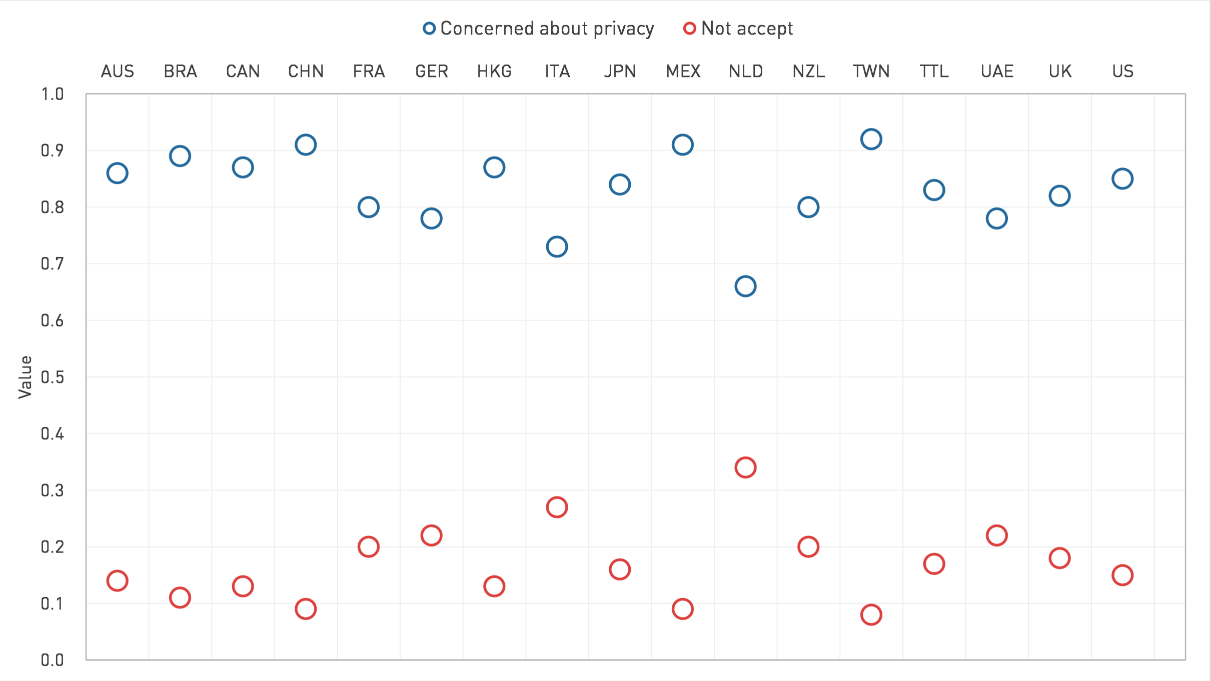
Chapter 2.
Decoding the Privacy Paradox: A Key Trade-off

Privacy protection is a fundamental requirement for an effective and vibrant society and economy. Some critical questions are: How much protection is enough? To what extent is privacy protection and personal information sharing a tradeoff? Are there effective principles, mechanism designs, and technologies that can protect privacy, even while promoting information exchange for the overall benefit of citizens and the broader society?

2.1 Privacy paradox: words or deeds?

In the digital economy, maximizing consumer and societal welfare depends increasingly on the effective analysis and use of personal customer data. What is each customer's willingness to disclose private information; how much do these customers care about protecting their privacy? Surprisingly, the answers to these questions, as revealed by survey responses and in actual behaviors, tend to go in opposite directions. This is known in the literature as the "privacy paradox": While most people say they care about privacy, they often appear to be too eager to share their information for free, or for littel compensation (Figure 1).

Figure 1. Privacy Attitude and Privacy Behavior



Source: Statista, 07/05/2019 and calculated by Luohan Academy

Note: The chart presents the gap between the percentage of global Internet users who claim to be concerned about their online privacy and the percentage of those who actually reject risks to their online privacy.

This seeming contradiction between people's presumed concerns over a fundamental right to privacy and their revealed behavior is a pattern that holds true throughout the world. Indeed, concerns about privacy have risen to become a global issue in the digital age, without regard to country or culture. According to the Eurobarometer data protection survey, about 67% of people in the European Union are concerned that they may not have control over the personal data they provide online. At the same time, a growing number of people recognize the need to provide such information.[3] A survey by Pew Research finds that 93% of U.S. adults believe that it is important to know who might have access to their information.[4] The China Consumers Association (2018) indicates that over 80% of survey respondents said that they have had their data leaked, and that they have received unsolicited pitches and advertisements. In a survey conducted by the Global Privacy Enforcement Network (2018) that involves 60 regulatory authorities in 39 jurisdictions, only 28% of the people surveyed had high trust and confidence in mobile, broadband, and utility providers, and only 15% in social messaging platforms.

Despite such widespread concern, many people are willing to relinquish their private data for small incentives. Throughout the world, online personal information sharing usually requires little or no direct financial reward. There is still less concern about sharing confidential information of others, if that is possible. For example, Athey et al. (2017) find that most participants in their experiment were willing to share their friends' email information in exchange for a pizza!

The privacy paradox has perplexed experts and policymakers. It is difficult to imagine that billions of people around the world would be irrational to such an extent when it comes to their private data, especially after so many openly express concerns about privacy. Even if some large Apps such as Facebook are hard to avoid, the same cannot be said of the majority of new Apps that are seemingly popping out of nowhere every day. The privacy paradox reveals a giant gap between "talking the talk" and "walking the walk." A proper understanding of this issue is important as it goes to the nature of the role of data and privacy and provides a crucial foundation for policy analysis.

Let us start then with a simple, commonly accepted definition of information or data privacy: "[C]ontrolling how one’s information is acquired and used" – what U.S. Supreme Court Justice Louis Brandeis called, "the right to be left alone" (Pavlou, 2011; Warren and Brandeis,1890). From this vantage point, we are able to examine how people make decisions when they must provide certain personal information in order to enjoy the benefits of mobile services.

There are several different “solutions” for the privacy paradox. Many argue that either the people involved are not informed or simply cannot understand the possible harm coming from privacy violations when they choose to share personal data, or that they simply have to tolerate certain privacy violations in order to use popular software applications (Apps) because of a lack of other options (Chen and Michael, 2012). But today many such options are emerging. An alternative, perhaps more plausible interpretation, holds that, when facing real-life decisions, it is people's actual behavior and not what they say about their privacy concerns that reveals how they truly perceive the tradeoff between privacy and data benefits. Both arguments remain hypotheses to be verified.

[3] See https://ec.europa.eu/commission/presscorner/detail/en/IP_15_6321.

[4] See https://www.pewresearch.org/internet/2015/05/20/americans-attitudes-about-privacy-securityand-surveillance/.

2.2 Digital information disclosure

The key is to understand users' behavior when they do have options, as they respond in the marketplace. We have conducted an empirical analysis using Alipay data. Alipay is a Chinese third-party mobile payment platform and enjoys more than one billion active users with millions of pop-up Apps called mini programs embedded in its platform. These mini programs serve businesses ranging from small startups to large corporations. Access to these Apps requires the user's permission to release certain personal data. Users can choose to opt out later by withdrawing permission. The Apps differ in the degree of sensitivity of the requested data and in the necessity of the requested data for their services. Combining these differences with user characteristics and choices, ***we provide the largest big-data study to date on privacy-related decisions.***

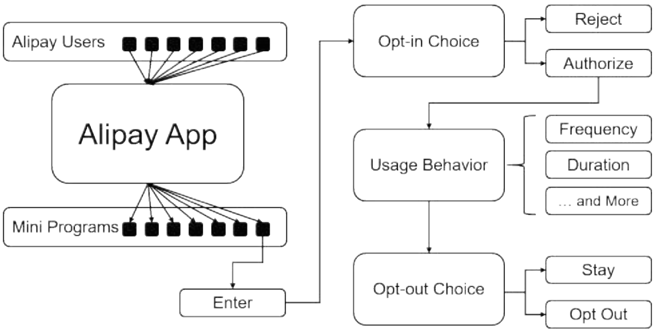
We focused on the opt-in and opt-out behavior of the mini program users, studied in Chen et al (2020). Studying users' choices on mini programs has several advantages. First, it provides a context in which users have to make actual decisions on whether they are willing to share information for services. In this case "talking the talk" is revealed as "walking the walk"— in the jargon of economists, "stated preferences" become "revealed preferences." Second, the one billion users include both sexes and they vary widely in age, income, education, and background. Accordingly, the data were used in one of the largest big-data field experiments on privacy-related choices ever conducted. Third, we utilized a sample with more than 50,000 mini programs that vary greatly in the benefits of the service and the extent of information required. While it can be argued that it is difficult for users not to use either Alipay or WeChat Pay, they do have plenty of discretion to decide whether they are willing to share some personal information for, say, subscribing to the services of a restaurant.

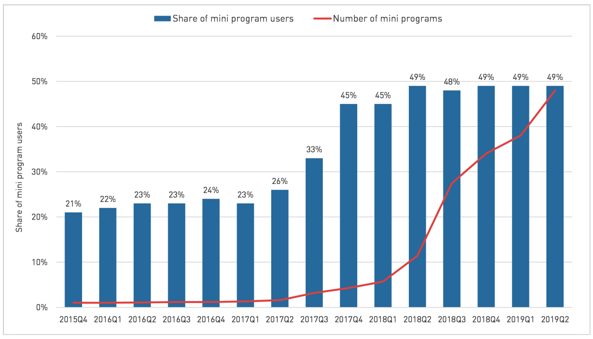
Poorly constructed surveys can fail to confront respondents with the full extent of such tradeoffs and difficulties. Anwyl (2011) explains, for example, why auto buyers often overstate their willingness to pay for fuel economy savings and carbon dioxide reductions, in part because they are not confronted with the costs they must incur, including the opportunity costs of, e.g., diminished carrying capacity, comfort, and safety.

Box 1 reveals many patterns regarding how different users make decisions on sharing personal information for obtaining services on the Alipay platform. These patterns are consistent with Chen et al. (2020), who provide much more in-depth detail on using the same setting with which to study and probably to better understand the privacy paradox.

Box 1. Mini programs on Alipay Platform

Alipay is a Chinese third-party mobile and online payment platform with more than one billion active users. The Alipay platform embeds millions of pop-up third-party applications called mini programs. Mini programs provide a myriad of services ranging from transportation to food takeout to financial services.





Information needs and sensitivity vary across the mini programs. Some mini programs simply request a user's account name, device location, and contact list, while others also demand the submission of personal credit scores. Combined with distinct user characteristics, such variations provide rich insights into the users' perceptions of the tradeoff between the need for personal data privacy and gains derived from the use of mini programs.

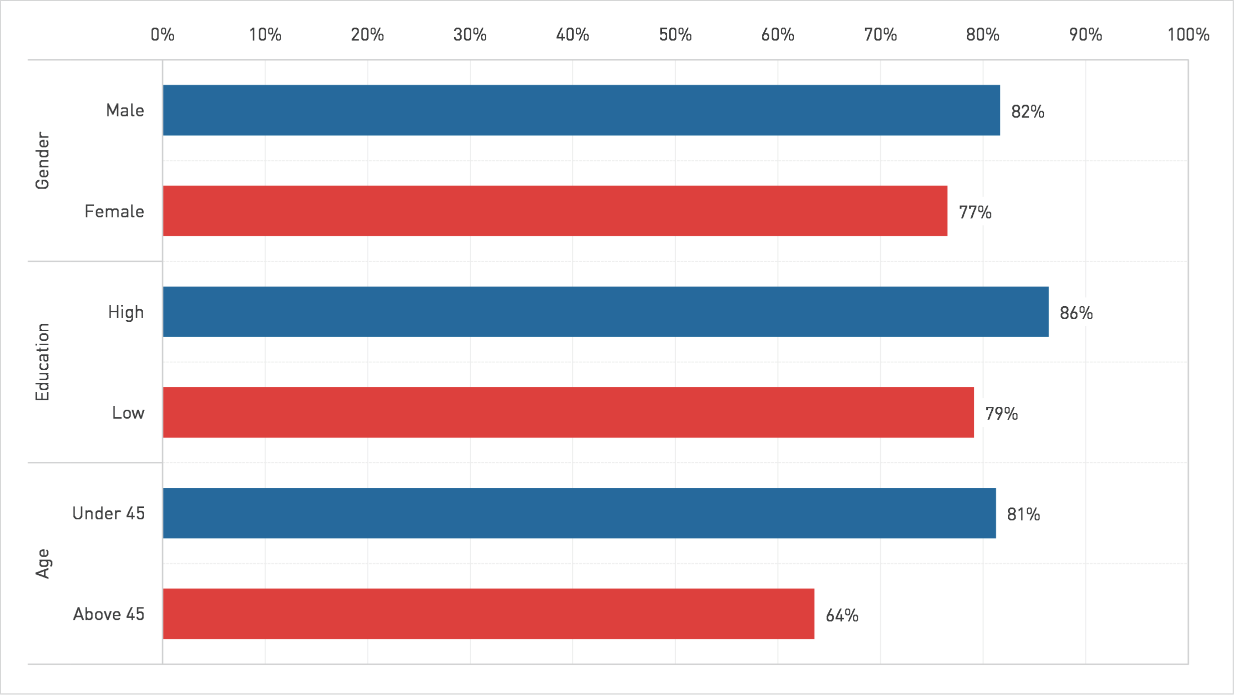
Millions of mini programs are embedded in Alipay, providing services including transportation, entertainment, government services, food and beverage, finance, and so forth. For example, a bike-sharing mini program may allow an individual to rent a bike without a deposit after being granted access to her credit scores. In recent years, the number of mini programs has soared on Alipay – from 2015 to 2019, the percentage of its users rose from 21% of total Alipay users to 49%, involving hundreds of millions of people.

Mini programs require prior authorization of data access from potential users. Such data may include a user's name, nicknames, gender, cell phone number, address, telephone broadband information, credit scores, payment account, transaction information and related services. Mini programs allow users to freely opt in or out.

What happens when users have options when choosing whether they are willing to share certain personal information in order to obtain services provided by various mini programs? As in other parts of the world, the average Chinese citizen is concerned about privacy. But when given options, she overwhelmingly chooses to share certain personal information in order to reap the benefits of new digital services.

The first pattern we have discovered in this study is that, when offered the opportunity, in the majority of cases, users tend to accept requests for information. As shown in Figure 2, the mini program opt-in rates for different types of consumers range between 64% and 86%. On average, over 75% of users choose to opt in. Male, higher-educated, and younger users are more willing to share their personal data in general. However, the difference is small across gender and education groups. People with college (or higher) education are more willing to accept mini programs, contrary to the hypothesis that people who are willing to share information tend to be uninformed. The largest discrepancy shows up among age groups. The percentage of people who are below 45 years old who choose to opt in is almost 20 percentage points higher than for people older than 45, perhaps partly because they tend to be more familiar with and better able to take advantage of online services.

Figure 2. Demographics of Opt-in Rates



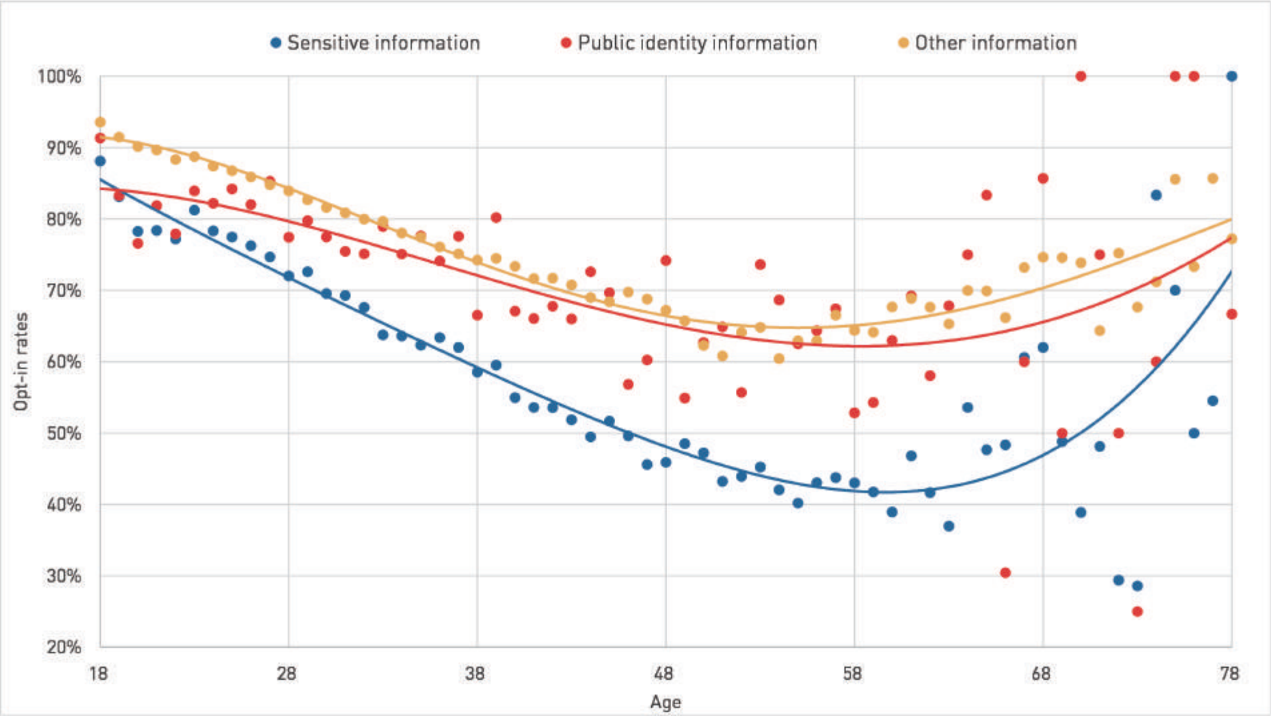
Source: *Luohan Academy*

Note: ***The opt-in rate is the share of the opt-in visits to mini programs that end up allowing the mini programs to access personal information. We define "highly educated" users as those with a bachelor's degree or above.***

People are not only willing to share personal information to join Apps in most cases, but they also rarely opt out, suggesting that they do not regret their initial decisions, at least not enough to take action. The overall opt-out rate turns out to be very low, about 1.2‰ per month during 2016-2019. This means that most users are willing to share their personal information in exchange for the gains from a valued service. They seldom change their earlier choices to opt in, either because they do not think it matters much or because they are content with their prior decisions. Johnson et al. (2020) found similar evidence that in 2010, only 0.23% of American users chose to opt out of targeted online advertising in a program called AdChoices. The rate was also low in Canada (0.16%) and in the European Union (0.26%).

The fact that users are willing to share personal information online does not mean that they do not care about privacy. ***Indeed, the more sensitive the personal data the fewer users are willing to share.*** Like Goldfarb and Tucker (2019) and Nissenbaum (2009), our study finds that privacy concerns are context-based: privacy concerns vary depending on the sensitivity of requested information (Figure 3). When mini programs require relatively more sensitive information, e.g., Alipay ID and car registration information, rather than general public identity information such as a nickname and profile image, the authorization rates decline on average by 20 percentage points. The gap is larger for older users, rising to over 30 percentage points for users between 55 and 65 years old. So, while people, in general, care about the sensitivity of their personal information, younger people are more open to sharing such information.

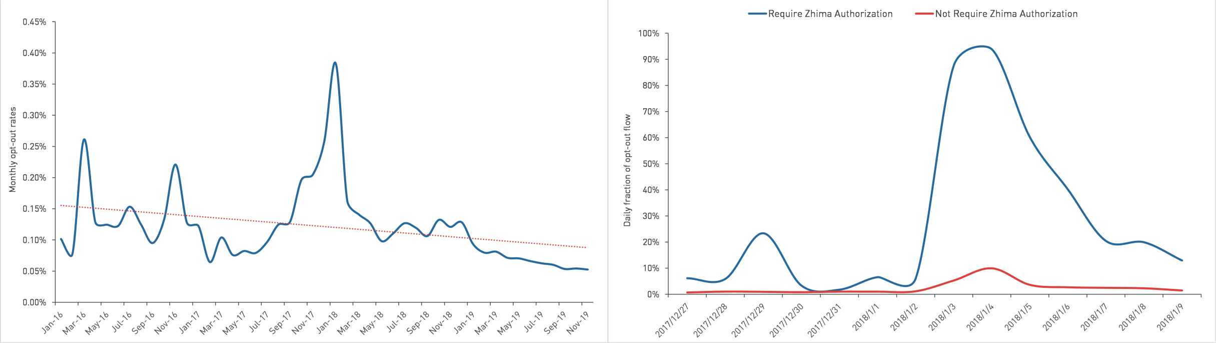
Figure 3. Opt-in Rates Over Information Requirements



Source: Luohan Academy

Importantly, opt-out choices are responsive to privacy incidents. This is revealed by a 2018 incident related to Alipay's Annual User Footprint Report, which led to a surge in the opt-out rates.[5] Alipay provides its customers annual reports that summarize their annual expenditures. Users usually are happy to share their spending experiences with friends. However, in January 2018, when customers opened their 2017 annual report in Alipay, there was a new clause asking for permission from users to opt in to the credit scoring service. The request was not prominent enough to be noticed, which raised concerns among some customers who believed that they should have been more explicitly informed of such changes. The opt-out rate rose sharply following the incident, from normally 0.12% per month to more than 0.3%, though it quickly returned to previous levels and has fallen further since late 2018 (Figure 4). Therefore, while users rarely opt out, they do respond when they feel that they have not been adequately informed of relevant changes in requests about their personal information.

Figure 4. Opt-out Rates Across Time



(a) ***Overall opt-out rates*** **(b)** ***More responsive for sensitive data***

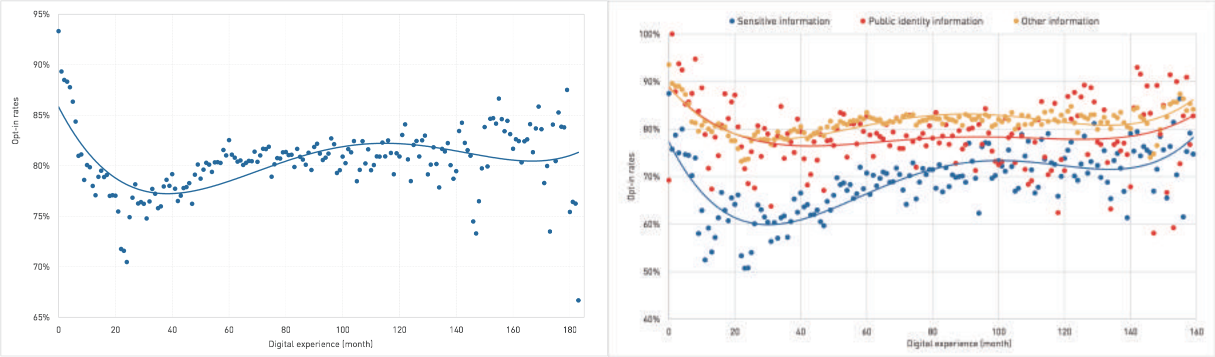
Source: Luohan Academy

Note: (a) The monthly average opt-out rate is calculated as number of opt-outs in each month over the accumulated number of opt-ins of mini programs. An opt-in happens when a user authorizes a mini program to access her information for the first time. An opt-out happens when a user purposely cancels the authorization of a mini program. (b) Daily fraction of opt-out flow refers to the proportion of daily opt-out users in the total of daily opt-out and opt-in users.

People's attitudes toward information sharing evolve with their digital experiences. Several previous studies found that as people learn more about new tools and technologies, they become more privacy-minded (Acquisti et al., 2015; Goldfarb and Tucker, 2012). Our study provides further, still more robust evidence in support of this hypothesis. The opt-in rate initially went down with the duration of Alipay use, suggesting that more experienced users are indeed more careful about information sharing than beginners (Figure 5a). However, for users with 40 months or more of digital experience, the opt-in rate rises to a level comparable to that for newer users. This pattern suggests a sort of U-shaped learning curve, with users increasingly embracing digital applications in the long run. Users' digital experience helps them to build knowledge and trust in dealing with personal information sharing. The same patterns hold regardless of information sensitivity (Figure 5b). Importantly, improved digital knowledge helps users embrace digital technology even more, with a greater willingness to share personal data.

[5] For more details, see https://www.scmp.com/tech/china-tech/article/2126772/chinas-ant-financial-apologises -over-alipay-user-data-gaffe.

Figure 5. Digital Experience, Information Requirement, and Opt-in Rates



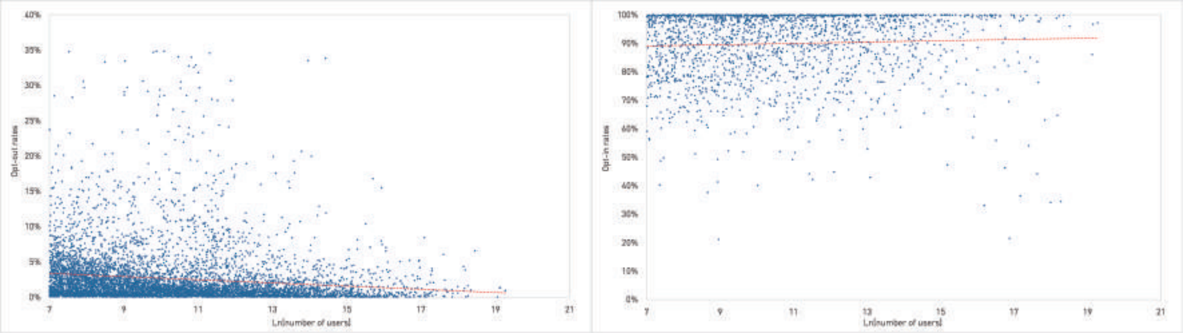
(a) ***opt-in rates and digital experience*** **(b)** ***opt-in rates and information requirements***

Source: Luohan Academy

Note: Digital experience is defined as the number of months since a user first registered Alipay.

Users are open to adopting new apps regardless of their initial popularity but unsurprisingly are more likely to opt out of less popular apps later (Figure 6). Here app popularity is measured by the number of current users. The flat relation between the number of users and the opt-in rate suggests that users are willing to try new services even when some services have few current users. If we regard app popularity as a proxy for the quality of the mini programs, the difference between the opt-in and opt-out behaviors could be explained by a better understanding of the mini programs' benefits after using them.

Figure 6. Opt-out Rates, Opt-in Rates, and App Popularity

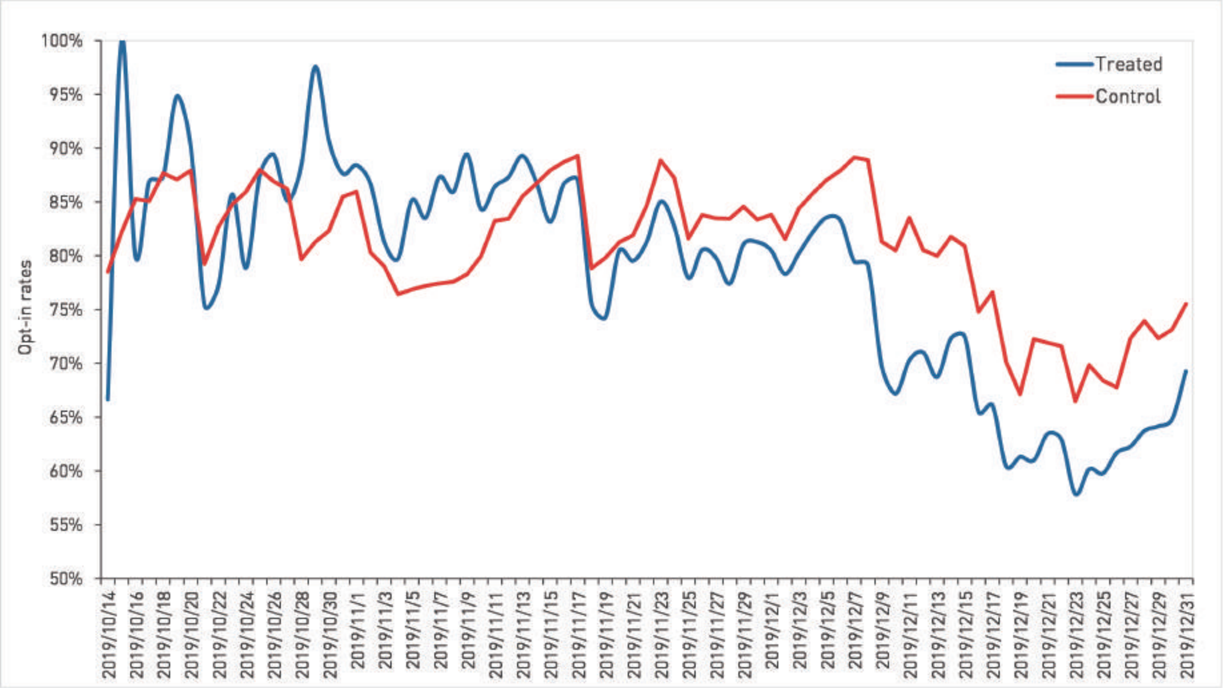


(a) ***opt-out rates and app popularity*** **(b)** ***opt-in rates and app popularity***

Source: Luohan Academy

Note: We define the popularity as the logarithm of users of each mini-program.

Figure 7. The Daily Opt-in Decision Before and After the Eperiment



Source: Luohan Academy

Lastly, users' trust in a well-known platform, in this case, the Alipay platform, makes them more willing to share their personal data. In an experiment conducted with Alipay mini programs in October 2019, the Alipay logo was removed from the user interface of mini programs. As a result, the opt-in rate fell by about 3% (Figure 7). The Alipay logo provided an implicit boost to trust in mini programs. Although the removal of the Alipay logo did not change the actual contracts between users and mini programs, it changed some users' perceptions.

A related example is Google's open-source Android Market, in which multiple parties can collect, transfer, and even transact data without much restriction, and users typically have less control over their personal data. According to the study by Kummer and Schulte (2019) involving 300,000 mobile applications, the reputation of application developers affects the privacy concerns of consumers while these applications are being installed.

In sum, ***Luohan Academy's study of "big data" reveals a world in which privacy concerns are real but are far from what they are sometimes thought to be in a hypothetical world in which users' privacy rights are summarily abused without choice.*** As in the rest of the world, the average Chinese citizen is concerned about privacy issues. But when they have a choice, the overwhelming pattern is that they are willing to share certain personal information in exchange for the benefits new digital services can give them. Both the degree of information sensitivity and the quality of services rendered play a role in their decisions. They do respond to privacy-related incidents. But otherwise, they seldom change their choices. The accumulation of digital experience helps them to be more prudent with their choices and ultimately leads them to embrace more digital services in the longer run. These patterns hold regardless of gender, age, or education.

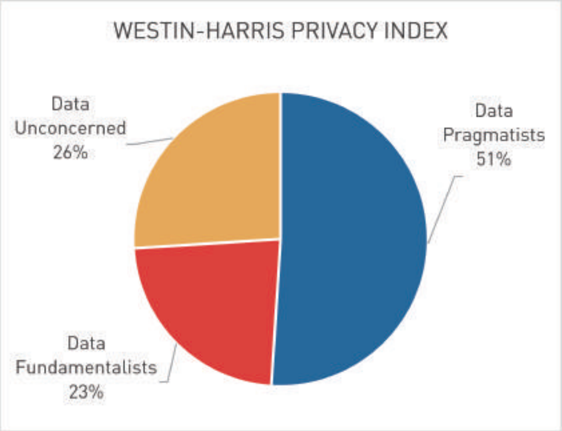
Clearly, privacy concerns are important, but they are only one consideration among others when users make decisions about whether to provide information sharing to obtain desired services. An integrated approach helps us to better understand the benefits and potential harms of data sharing.

Finally, and even more significantly, Chen et al. (2020) find that people who are more concerned about the "privacy paradox" are the ones most likely to adopt the mini programs that address their concerns. This implies that users are concerned about a potential "paradox" precisely because they are willing to share personal data to obtain services but, in the meantime, have a demand for relevant privacy and data security. The fact that more concerned people tend to share information more also suggests that this is not an issue of ignorance or irrationality. Rather, it highlights the value of data sharing, and the need to find better and cheaper ways to protect personal privacy. Put differently, if the overwhelming pattern is users' willingness to share personal data in order to enjoy services, then the best policy to deal with personal data is not to lock it up or to make it costly to share, but to promote privacy and data security more effectively and efficiently, allowing consumers to make tradeoffs that once they were unable to make.

2.3 Weighing the risks of personal information disclosure

Different respondents can provide very different insights on how users perceive personal information sharing. By investigating individuals' perceptions, distrust, and fears over information technology, the Westin-Harris consumer privacy surveys from the late 1970s to 2004 produced a privacy index that classified individuals into three categories: “privacy fundamentalists” -- those who are unwilling to provide personal information despite the prospect for service enhancements, “privacy pragmatists” -- those who make decisions on a case-by-case basis based on whether a service or service-enhancement offered is worth the information requested, and the “unconcerned” -- those with few worries about the collection and use of their personal information (each of whom may have had rational reasons for making tradeoffs given their personal preferences for security versus the information received on the Internet at the time) (Equifax-Harris Consumer Privacy Survey, 1991, see Figure 8).

Figure 8. Consumer’s Willingness to Disclose Personal Information

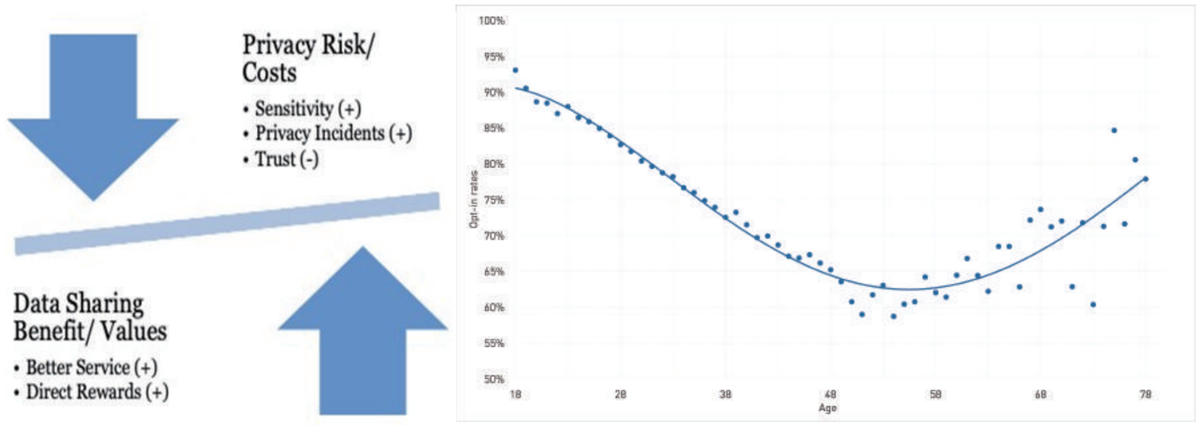


Source: GDMA (2018)

More recently, by analyzing consumer responses across 10 global markets, the Global Alliance of Data-Driven Marketing Associations (GDMA)[6] has found that 51% of the population across these global markets are ***pragmatists,*** 23% are ***fundamentalists,*** and 26% are ***unconcerned.*** These findings and our analysis lead us to conclude that most users are willing to share, and indeed do share personal information in exchange for the benefits of offered services; this despite concerns about the sensitivity of their data.

By approaching privacy-related issues from the perspective of decision-making, we can further understand the value consumers attach to privacy as well as user preferences regarding the sharing of their personal information. Motivated by the work of Kahneman and Tversky (1984), a consensus has emerged that consumers can experience both benefits and risks when disclosing their personal information (Culnan and Bies, 2003). This consensus later developed into the so-called "privacy calculus" (Culnan and Bies, 2003; Dinev and Hart, 2006; Laufer and Wolfe, 1977; H. J. Smith et al., 2011; Xu et al., 2009). The privacy calculus states that consumers generally perform a risk-benefit analysis before disclosing data, analyzing the risks of sharing their personal data relative to the benefits, as illustrated in Figure 9a. Consumers, thinking this way, are predicted to disclose their personal information so long as the perceived benefits outweigh the perceived risks, and vice versa.

Figure 9. Privacy Calculus and Consumer Opt-in Rates



(a) ***Privacy calculus*** **(b)** ***opt-in rates and consumers' age***

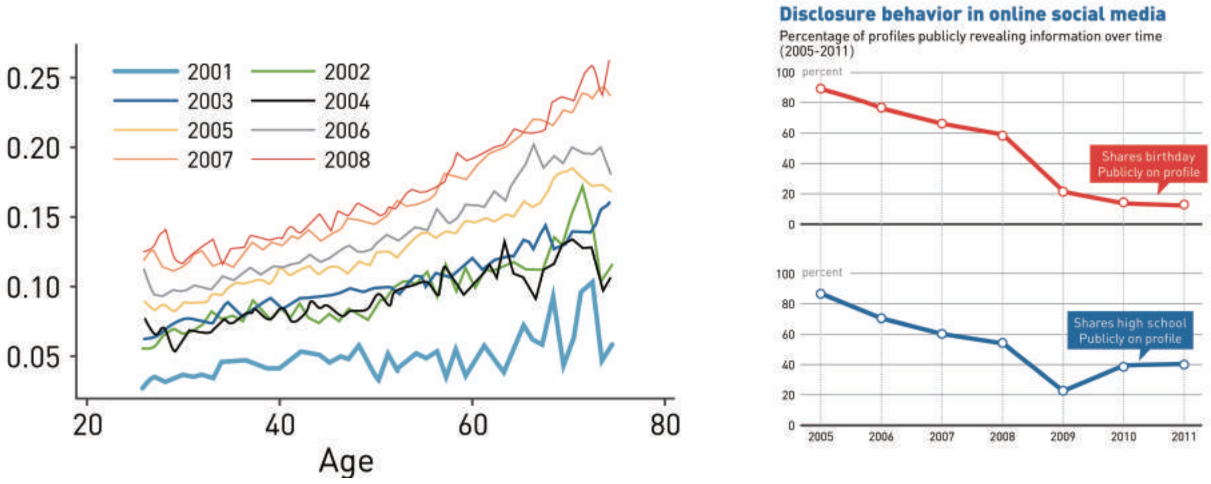
Source: Luohan Academy

[6] See GDMA (2018) Global Data Privacy: What the consumer really thinks?

The evidence from Alipay's mini program users (Figure 9) is consistent with the privacy calculus perspective. Information sensitivity, privacy incidents, and the quality of services all matter when users make information disclosure decisions. Importantly, different people, or even the same people with evolving digital exposure, place different weights on these considerations. Young people embrace digital services most enthusiastically, presumably because they have less to hide, are more care-free, have a better familiarity with and enjoy the benefits of information exchange more than other age groups (Figure 9b). Interestingly, the opt-in rate goes down monotonically from the 20-year-old age group to the 50-year-old age group and then climbs again. This might reflect the fact that the 50-year-old group coincides with the retirement age in China. This group has accumulated more wealth and social relations and responsibilities than other groups.

Past work suggests that privacy concerns are trending up. Goldfarb and Tucker (2012) show rising privacy concerns in the United States (Figure 10a). They measure how likely individuals are willing to reveal their income information via an online survey from 2001 to 2008. They find that the refusal probability increased by 1.3% per year. This trend holds for all age groups. Similarly, Acquisti et al. (2015) find that the overall percentage of members within the Carnegie Mellon University Facebook network, who chose to share their birthday and high school information publicly, declined from 2005 to 2011 (Figure 10b).

Figure 10. Disclosure Behavior Over Time



(a) ***Percentage refusing to reveal income information***

(b) Percentage revealing information in online social media

Source: (a) Goldfarb and Tucker (2012); (b) Acquisti et al. (2015)

Chapter 3.
The Value of Data

The fact that a majority of users are willing to share personal information with providers of services reveals their desire to enjoy the benefits of digital services. This naturally raises the question of where the value of data sharing lies. Just how valuable can the information be? How is it used? What has been changed to make it so important in the digital era? In this chapter we seek to answer these questions by looking at the value of online data sharing from three perspectives: ***connectivity, decision making, and trust.***

3.1 The transformed importance of information in the digital age

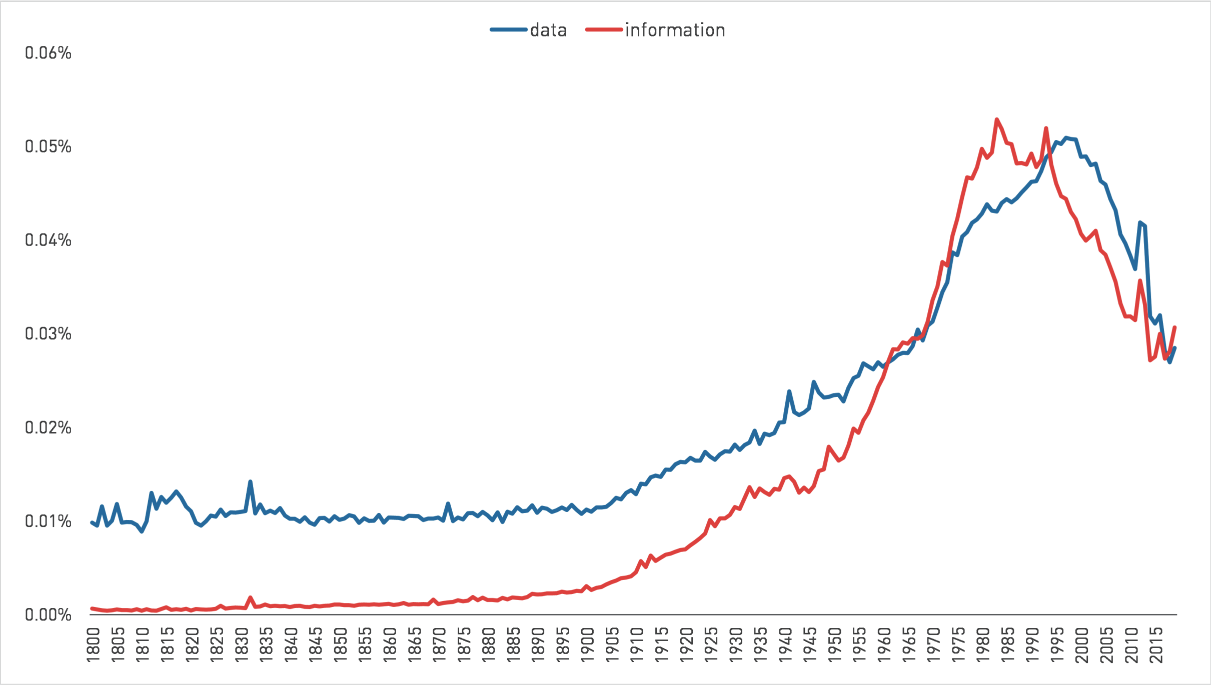
Economists have long realized that the exchange of information is indispensable for economic activity. Hayek (1945) made two critical points in this regard. First, unlike a normal commodity, the information required to make decisions frequently does not exist in concrete or integrated form. Rather, it needs to be produced and processed. Second, to be able to make use of all the dispersed information， ***the*** economic problem of society is to facilitate information collection and exchange.

This may sound like simple common sense. But in the increasingly heated debate about privacy concerns, the focus is frequently on the downside of personal information exchange. In part, it reflects the increasing harm done by privacy breaches, identity theft, and cybercrimes in the digital age. Yet, we should not forget that information sharing is essential, not only to the welfare of each individual, but also to the progress of society as a whole.

Path-breaking economic studies in the 1970s and 1980s showed how limited and asymmetric information can be a major obstacle to the efficient allocation of goods and services, suppressing otherwise voluntary and mutually beneficial trade, giving rise to market failures that impair the effectiveness of macroeconomic policies, distorting investment and consumption decisions, and generating unemployment that would otherwise not exist in a competitive equilibrium (Akerlof, 1970; Milgrom and Stokey, 1982; Myerson and Satterthwaite, 1983; Phelps et al., 1970; and Pissarides, 2000). The sharing and distribution of information affects the degree of human cooperation. Clever market design and mechanisms can to some extent mitigate the distortions arising from asymmetric information. These include signaling (Spence, 1973), screening (Vickrey, 1961; Mirrlees, 1971), as well as more general mechanism designs (Dasgupta, Hammond and Maskin, 1979; Green and Laffont, 1981; Myerson, 1981; Maskin, 1983, 1999, 2008) and matching protocols.

The information revolution began in the 1940s, fundamentally transforming the way we produce and use information. In 1946, the word "data" was first used to mean "transmissible and storable computer information." [7] As shown in Figure 11, by the 1970s, the use of the term "data" had surpassed the use of the term "information," suggesting a rising awareness of the potential for the use of digitized "data" as the information age forged ahead (Figure 11).

Figure 11. Reference to "Information" and "Data" in Published Books, 1800-2004



Source: the Google Books Ngram Viewer

Note: This graph shows the frequencies of the terms "data" and "information", in unigrams or number of words, in google books published in English, from 1800 to 2019.

The explosive use of data has been driven by the unprecedented decline in the cost of data formation, production, storage, and communication. This trend is illustrated by ***Moore's Law,*** the empirical observation (not really a "law") that the number of transistors that can be stored on a microchip and used in a dense integrated circuit doubles about every two years, accompanied by a similar decline in computing costs. That pattern led to similar advances in digital electronics, such as the reduction of quality-adjusted microprocessor prices, the increase in memory capacity, and the improvement of sensors. Advances in other computing fields have further accelerated the growth in data production. For instance, cloud computing provides shareable data processing capacities that allow for more efficient business computation. New technologies such as silicon photonics, which uses light to transfer data, show great promise to further accelerate the processing of data. Advances in artificial intelligence and machine learning point to still further, ever-increasing improvements in the processing and use of pieces of data, though perhaps not as rapidly as in the past.

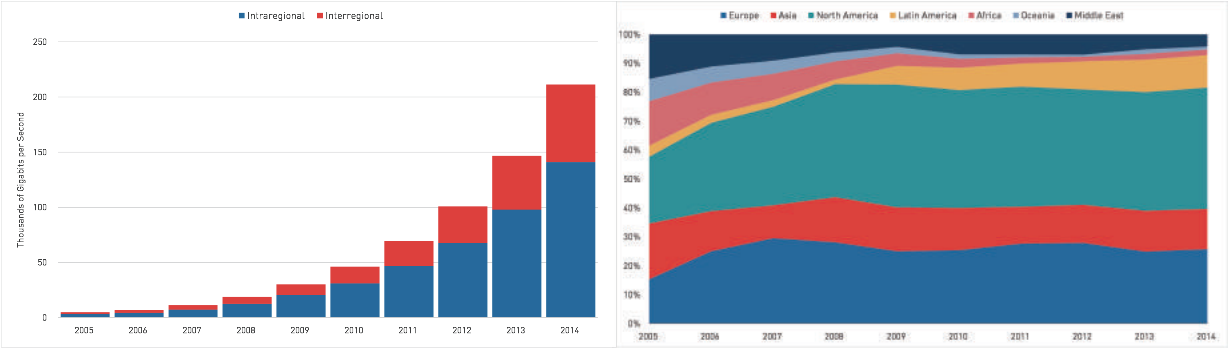
[7] See "data | Origin and meaning of data by Online Etymology Dictionary" at www.etymonline.com.

Though significant startup costs remain, the digitization of information has advanced to the point where the marginal (incremental) cost of storing and transmitting increasing volumes of data has fallen to virtually zero -- to the point where the use of the term "data" is more ubiquitous than that of "information." Further, the term "data" has morphed into "big data" characterized by the well-known three V's: ***Volume, Variety,*** and ***Velocity,*** which, as noted in Chapter 1, capture the unprecedented scale, usefulness, and speed of data processing.

Indeed, the total volume of data created and stored rose from a mere 0.8 Zettabyte (ZB) or trillion gigabytes in 2009 to 33 Zettabytes ZBs in 2018. And it is expected to reach 175 ZB in 2025 (Reinsel et al., 2018). As for variety, literally all kinds of information can be digitalized as data because it is now so cost-efficient. The third V, velocity, has also increased. Data are now generated at and flow at once-unimaginable speed. For example, as a message is tweeted, it is recorded in Twitter's data architecture and microseconds later it is published onto the user timelines. Thousands of millions of Twitter users all over the world tweet, resulting in massive and real-time data flows 24/7. What "velocity" in big data means is an unprecedented amount and breadth of the production and sharing of information in real-time.

The pace of data flows continues to accelerate, with the speed of global internet traffic from 2,000 Gigabytes per second in 2007 to 46,000 GB per second in 2017 (UNCTAD, 2019), a more than 20-fold increase in 10 years. This is equivalent to about four times the size of the entire U.S. Library of Congress collection being transferred per second! It has become a global phenomenon, as data flows, both within and across regions, have increased exponentially around the world. (Figure 12)

Figure 12. Data Flows Rose Exponentially Rround the World

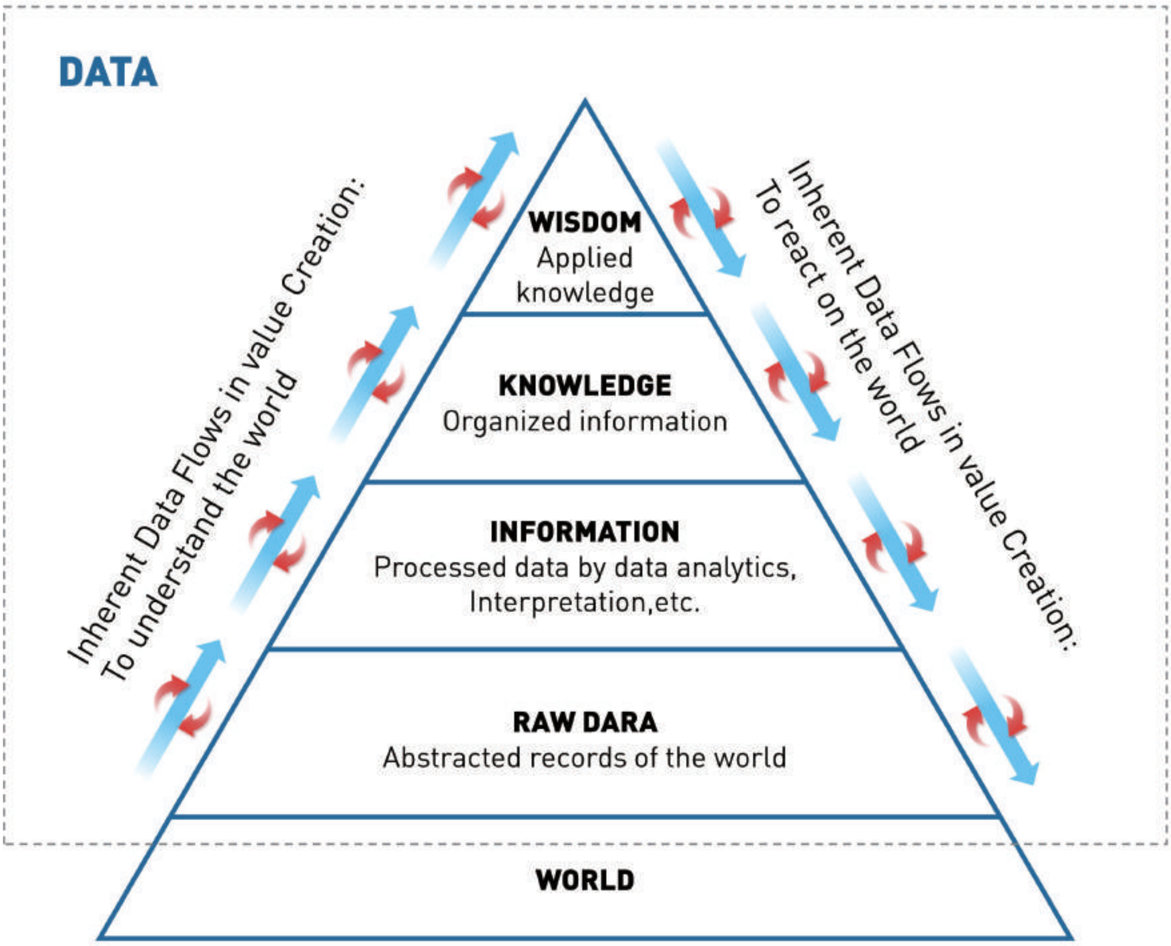


(a) ***interregional and intraregional data flow increase*** **(b)** ***regional data flows***

Source: McKinsey (2016)

As we also noted in the introduction, ***"data" is not the same as "information."*** "Data," defined as a digitalized form of records, can be described as a carrier or agent of information. But data do not necessarily contain information. For instance, adding a set of data generated by a random walk does not convey additional information. In addition, it takes effort to produce (observe) raw data from the physical or virtual world and process the data to gain information about its economic or social value (Figure 13). In addition, as noted by Singh (1999), by developing algorithms that bury information deeply in data, cryptography reveals and exploits the big difference between data and information. While data itself can be made "public" and hence freely available, only those in possession of the encryption "key" are able to extract information from it. Conscious of this difference, and as noted in the introductory chapter, we use "information" and "data" interchangeably and point out their difference only when necessary.

Figure 13. Knowledge Pyramid (adapted from Kitchin (2014) and Boisot and Canals (2004))



Source: Luohan Academy

3.2 The value of data in the digital age

Data are only valuable when they are in use, which is exactly what is going in nearly every corner of the world. The information revolution is leading to fundamental changes in ***connectivity, decision making, and trust,*** the three building blocks for coordination. The following discussion examines each of these in detail.

(a) Digitized connectivity: An unprecedented level of inclusive participation and coordination

As we demonstrated` in ***Digital Technology and Inclusive Growth*** (Luohan Academy, 2019), ***data sharing enhances connectivity.*** Because digital data are so easy to produce and share, there has arisen an unprecedented level of ***inclusive connectivity*** that has reshaped the marketplace and the way people coordinate their business and consumer activities.

The greater affordability of producing and exchanging data has led to record levels of participation in information sharing activities. Eight out of ten adults in the developing world now have mobile phones. There are now more people with access to mobile phones than to clean water (World Bank Group, 2016). The fact that the world is increasingly interconnected through digital technology is redefining markets and the way people coordinate the production and allocation of goods and services throughout the world.

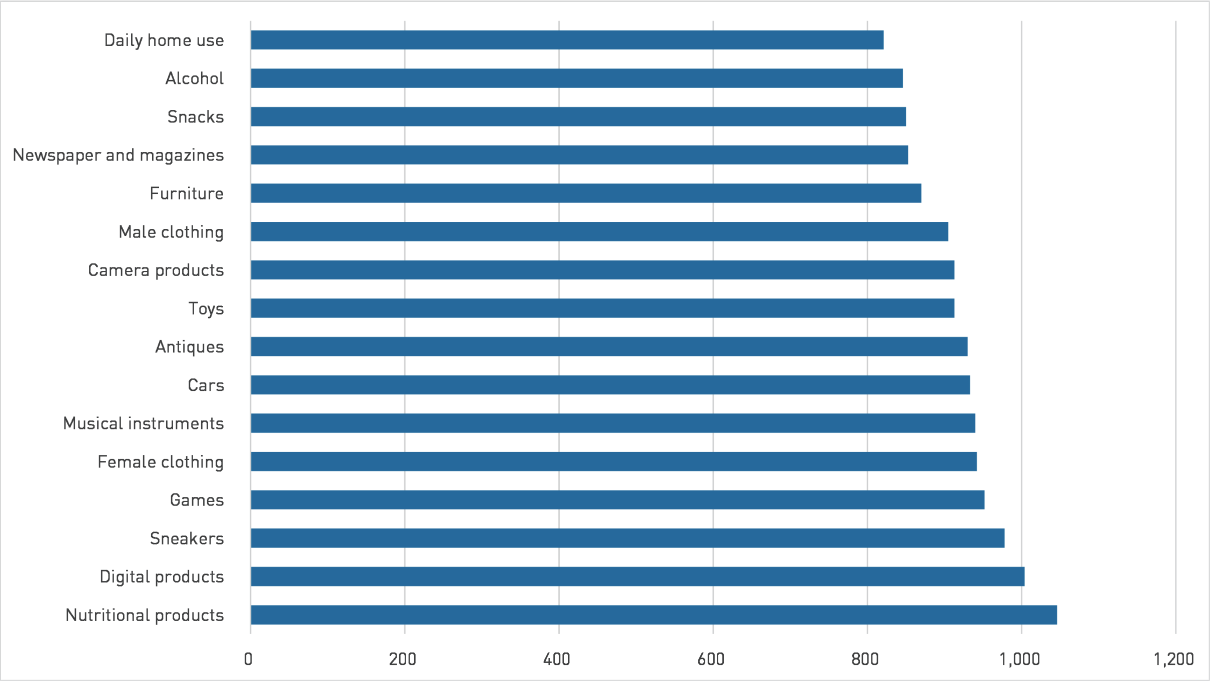
In ***The Wealth of Nations,*** Adam Smith observed that "the division of labor is limited by the extent of the market." Trade has long been universally defined by the "gravity model"- the amount and frequency of trade are negatively correlated with the distance between agents, and positively correlated with the size of markets. Digital commerce has dramatically expanded the scope and depth of markets -- well beyond what Smith and others of his time ever could have imagined.

Consider Shanghai, one of the most cosmopolitan cities in China. Even in the most popular business districts of Shanghai, more than 80 percent of offline customers live within 10 kilometers of the center of the trading area.[8] At longer distances, buyers and sellers simply do not know each other. They lack accurate information on the variety, quality, and prices of commodities and services, as well as on the details of customer demand and seller reputation.

Across the world, two-sided e-commerce platforms such as Alibaba and eBay give a picture that would have defied Smith’s imagination. Online market exchange has facilitated an enormous increase in the scope, depth, and breadth of trade. Offline trade has long been universally described by the "gravity model," in which the majority of customers in a local market come from within a 10-kilometer radius. The picture on Alibaba’s Taobao App for online trade stands in stark contrast. Each month more than 720 million active users shop there. They are served by more than ten million startups and companies. Half of these entrepreneurs on Taobao are women. Over three billion listed commodities and services can be ordered online. Excluding fresh food, the average shopping distance between a buyer and seller is close to 1,000 kilometers, two orders of magnitude larger than the historical norm. The shackles of the gravity model on trade have been broken.

[8]. See the report from MetroData Tech, Shanghai. http://www.sifl.org.cn/show.asp?id=3956

Figure 14. Average Trading Distance by Product Category on Taobao and Tmall, 2018



Source: Luohan Academy (2019)

Note: The sample consists of all transactions made on a randomly selected day in 2018. Distance is calculated as the diametric distance between the capital cities of the sending and the receiving province.

What makes this new form of the market possible is the flow of information and the matching of consumers and producers. With billions of items of commodities and services available, it is simply impossible for customers to go through all the potential items of interest; nor could producers reach all potential customers. If a key obstacle that restricts traditional markets was the lack of information, then the new obstacle in the digital age is information overload. What is needed is not data or mere information, but relevant information that is valuable to the user. An efficient matching mechanism based on big data from both consumers and producers is therefore critical.

(b) Data sharing enhances decision-making

Large volumes and varieties of data, combined with connectivity, are enabling countless customers and producers to make smarter decisions, leading to more rapid and more beneficial product innovations, new and more innovative sales and services, and new and more innovative business models—new methods of industrial organization—that were simply not possible before (Luohan Academy, 2019) (See also the discussion of Schumpeterian competition in Chapter 6).

E-commerce platforms increasingly use recommendation systems to better help consumers find the products they want, in addition, of course, to traditional tools such as search and shop listings. Digital platforms use consumer data, such as purchasing history, search activities, and personal attributes as inputs, to predict what goods and services the consumer most likely needs. The matching recommendations are done through algorithms so that suppliers "know their customers but do not know who they are." As effective as these matching algorithms can be, buyers and sellers are only now beginning to explore the potential for the technology. And so far only a tiny amount of the relevant data is being applied to help match buyers and sellers, users and suppliers.

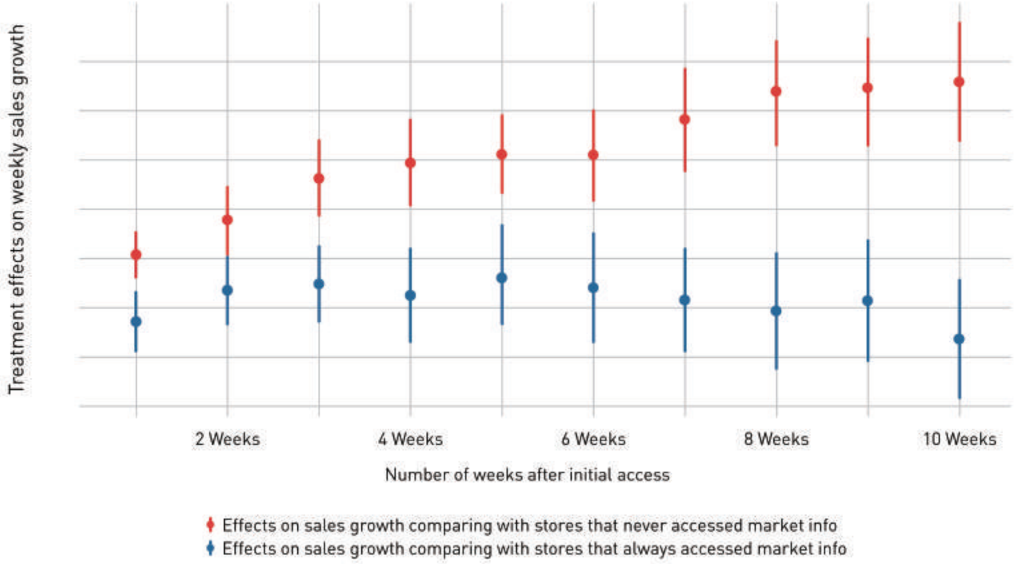
The enhanced flow of data not only helps consumers make smarter shopping decisions. It also helps producers better understand their customers and make smarter customer-oriented decisions.[9] This is particularly relevant for micro, small, and medium-sized enterprises (MSMEs) and individuals that, until now, have enjoyed little or no access to much-needed product and customer information. One example is Alibaba's Business Advisor, a service available to all online store owners for various information-analytic tools, such as sellers' historical performance, market trends, and the nature of their competitors. New subscribers, most of whom are MSMEs, typically experience a significant jump in sales growth within the first week of subscription, and the difference in performance between the subscribed and unsubscribed groups increases steadily over the next ten weeks. In addition, the data sharing made possible by "big data" also helps MSMEs to grow by equipping them with a wide array of sophisticated analytical tools that used to be available only to large corporations.

Digital platforms such as Taobao, JD.com, Amazon, and eBay all provide information services to help sellers, regardless of their size, to make business decisions. Consider Taobao's Business Advisor, a service available to all online store owners. It provides business intelligence, from free basic packages to paid services, helping firms make more informed decisions. It provides business owners with useful analyses, such as those of their historical performance, market trends, and market competition. According to Luohan Academy (2019), about 90% of online sellers with monthly sales greater than CNY 300,000 (USD 44,000) are subscribers to the Business Advisor. Access to these data allows subscribers to adjust their business strategies to ever-changing, often unpredictable circumstances. These include price adjustments, product listings on online stores, and more frequent changes to the titles of listed products.

New subscribers to the Business Advisor, most of whom are SMEs, immediately benefit from the availability of customer and market information. Becoming an active user of Business Advisor, i.e., starting to access data on a store's own performance, is associated with approximately 15% higher average weekly sales. One extra day of tracking per week is related to 6.9% higher annual sales. New subscribers significantly outperform non-users. Their weekly sales are about 13% higher in the tenth week after the start of subscriptions. In short, big data helps SMEs gain business analytical tools that used to be the privilege of a few large corporations who could afford the expense.

[9] See Veldkamp and Chung (2019), Carriere-Swallow and Haksar (2019), and Jones and Tonetti (2020), for a summary of recent works about the role of data in the aggregate economy by providing new information.

Figure 15. The Effects of Acquiring Market Information on Weekly Ales Growth



Source: Luohan Academy (2019)

Note: Compared to stores that have never accessed market information, stores that recently acquired such information experience sizable subsequent growth in sales. The growth in sales of recent users compared to stores that always have access to market information is still positive but not as strong.

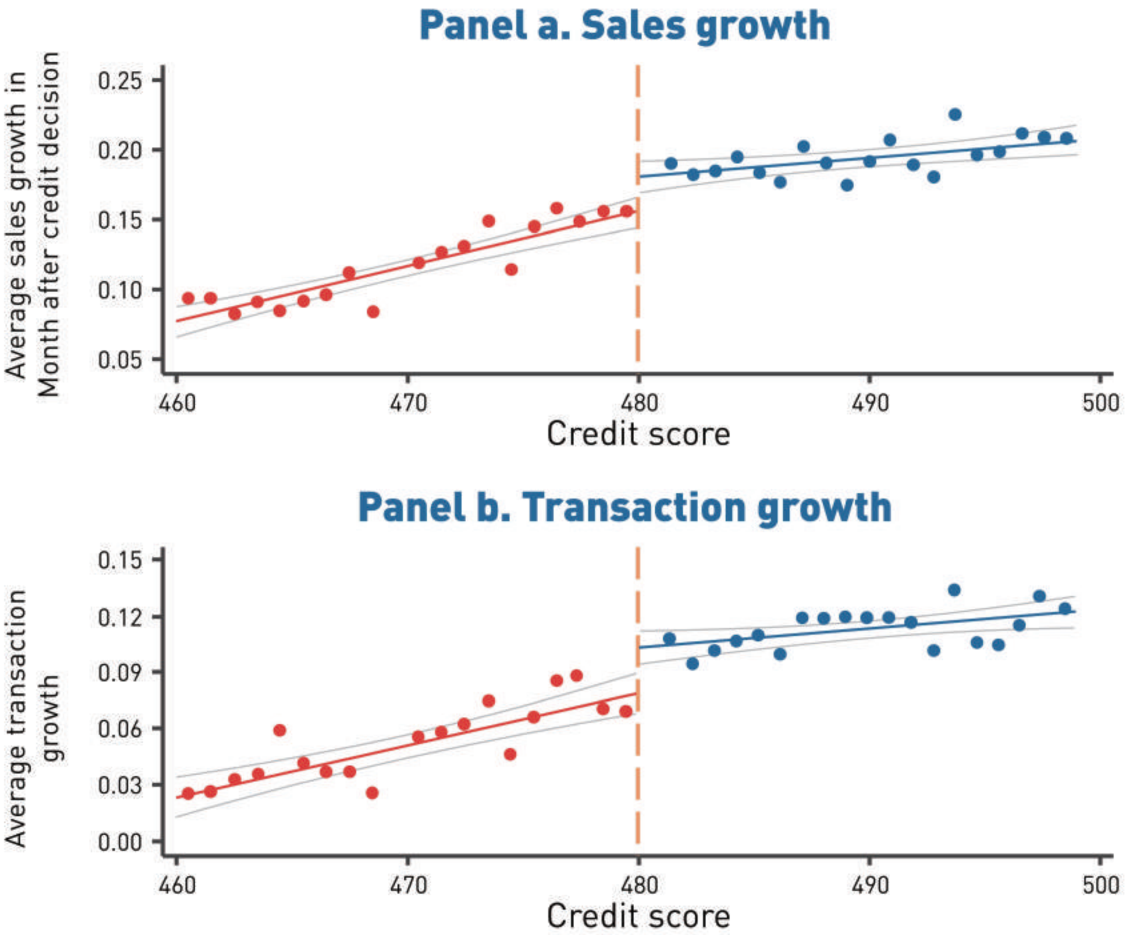
Small businesses also benefit from the availability of finance without the need to commit physical collateral, solving a heretofore largely insurmountable barrier to inclusive finance. This has opened up opportunities for tens of millions of Chinese and Kenyan entrepreneurs – a central finding of Luohan Academy's 2019 report on inclusivity. Big data has enabled the emergence of new large-scale microloans that were simply not possible before.

Finance is another area where “KYC” (Know-Your-Customer) is extremely important. Historically, the majority of business lending has been collateral-based and most MSMEs (micro, small, and medium enterprises) have not been able to borrow because they lacked collateral. This perpetuated a huge shortage of MSME funding, miring tens of millions of Chinese entrepreneurs and hundreds of millions of their potential hires in poverty.

Fintech has flipped the script. By reducing the need for collateral, the emergence of fintech has contributed to promoting financial inclusion, providing credit to large volumes of heretofore unserved or underserved firms in the MSME universe (Hau et al., 2018). Fintech lending uses big data to serve millions of SMEs with low collateral but high growth potential, unlike traditional financial intermediaries who rely on information-insensitive collateralized loans, mostly to larger corporations. Since 2011, MyBank has lent to more than 20 million SMEs and startups without collateral. It uses a "310" model: It takes an MSME less than three minutes to apply, one second to obtain the loan, and zero personnel to complete the transaction. Information available about the borrower is enough to assure the lender that it is a risk worth taking (Luohan Academy, 2019).

Thanks to big data and the digitization of finance, information has become the new "collateral" that has helped make many a budding entrepreneur successful (Holmström, 2018). Using a lending decision rule of a 480-point credit score, Hau et al. (2018) show that MSMEs that obtain lending enjoy significantly higher sales growth than those who do not enjoy such access.

Figure 16. Sales Growth After Credit Approval



Source: Hau et al. (2018)

Note: The data sample period is from September 2014 to July 2016.

Similarly, Berg et al. (2020) have studied the effects of digital footprints on credit evaluations. “Digital footprints” are the type of information consumers leave online when registering or browsing a website. The authors find that even simple digital footprint-information can become a useful supplement to current credit bureau information. When using a combination of both credit bureau and digital footprint-information, banks have been able to reduce default rates by roughly one-third. Such information can also increase access to credit for unbanked populations.

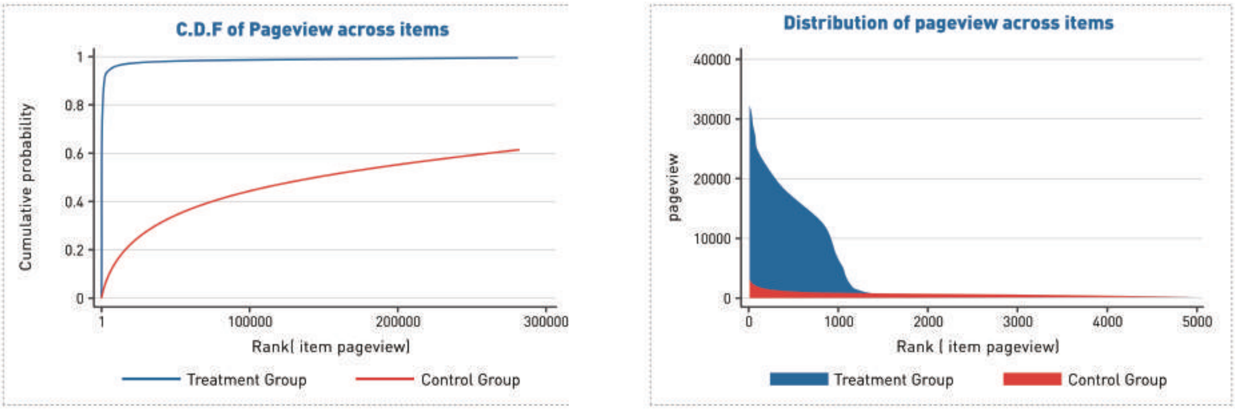
Aside from a historical lack of access to lending, a lack of financial accounts has always been a widespread problem for billions of impoverished people globally. Strikingly, within the space of ten years, China has transformed itself into a mobile payment-based country with more than one billion users, many of whom have been lifted out of poverty.

What has made this widespread digital payment system reliable and sustainable is the use of big data with its "three Vs," expanding access to large volumes of borrowers with diverse needs at rapid speed. When someone makes a payment by swiping a traditional physical credit or debit card, there is little knowledge of who is doing this. In contrast, with mobile payments, there is additional information such as location, time, biometric features, consumption habits, and identity of buyers and sellers. This additional information allows a mobile payment algorithm to assess whether the person making the payment owns the account, whether there is enough money in the account, and whether the person is making payment decisions that are unusual (a red flag). According to the People's Bank of China, the average fraud loss rate on credit cards and debit cards is 0.02%. But the fraud loss rate on typical mobile payments in China, using Alipay as an example, is less than 0.0005%. Users, including the unbanked population, can now enjoy mobile payments backed by real-time risk assessment, which in turn depends crucially on the real-time exchange of actual user data.

How important is the use of personal data in this process? What would happen if the recommendation is done without using personal data? To address these issues, Sun et al. (2020) carried out a large-scale randomized field experiment involving more than 620,000 users. Users normally see a panel of recommended products when they visit Taobao, a retail e-commerce platform of Alibaba’s. Recommendations are personalized using an automated matching algorithm that uses relevant personal data as part of the inputs. In the experiment, all participants were randomly assigned to either the treatment group or the control group. The matching algorithm for product recommendations stayed the same for the control group, but the use of personal data was shut down for the treatment group. This comparison between the treatment group and the control group helped quantify the value of personal data.

The outcome was striking. With personal data being used in the matching, customer page views were roughly evenly distributed across top items and long tails (Figure 17). Without personal data, recommendations became significantly concentrated in a small number of products, with little variety, and the number of effective matches between the recommended products and customers also dropped. The top 1,000 items received almost 90% of all exposures. Schumpeter's gale force of competition was reduced to a light breeze.

Figure 17. Average Trading Distance by Product Category on Taobao and Tmall, 2018

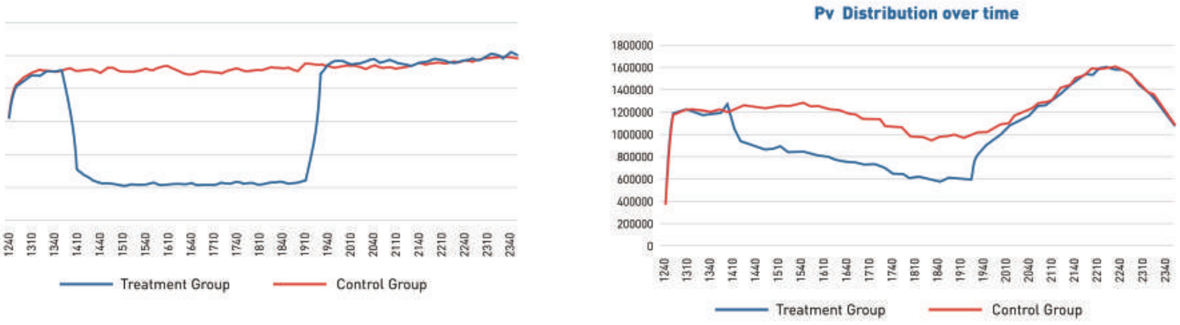


Source: *Sun et al. (2020)*

Note: ***Without the use of personal data, the recommendations concentrate on a small subset of items and cover a much lesser variety of items. Top 1000 items receive almost 90% of all exposures. "C.D.F." means "cumulative distribution function."***
["Pageview" should be plural in both charts.]

Along with a drop in the variety of available products, consumers became much less engaged. The lack of personal data led to a drop of 77% in the click-through rates and the number of product views fell by 33% (Figure 18). As the products shown to consumers were not so attractive, the number of searches increased significantly.

Figure 18. User Click-Through Rates and User Product Views



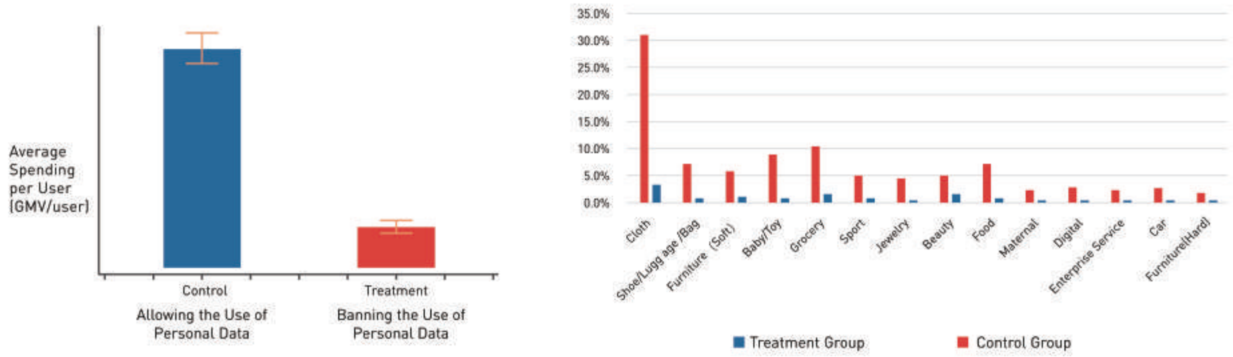
(a) ***User click-through rates on recommended itemsdrops by 77% (Left panel).***

(b) ***User product views on homepage recommendation drop by 33% (Right panel)***

Source: *Sun et al. (2020)*

As a result of ineffective matching and customer dissatisfaction, transactions declined by a whopping 81%. So, a reduction in the listing of recommended products and less active customer engagement led to a significant decrease in the volume and value of market transactions (Figure 19). In fact, an increasingly large part of online consumption nowadays relies heavily on personalized recommendations. The power of the automated recommendation system would be greatly weakened without the input of personal data. It is only when the suppliers "know" their customers that customers' demands can be satisfied. This is exactly the problem when privacy regulation becomes too stringent – when the risks avoided are not justified by the loss of valuable recommended data, as judged by consumers themselves.

Figure 19. Transactions by the Control and Treatment Groups



(a) ***Gross Merchandise Volume (GMV) drops by 81% and the Number of transactions drops by 86% (Left Panel);***

(b) ***The Impact of Privacy Regulation is Large across All Industries/Categories in E-commerce (Right Panel). Source: Sun et al. (2020).***

Source: *Sun et al. (2020)*

Importantly, and contrary to concerns that "companies could use big data to exclude low-income and underserved communities from credit and employment opportunities" (FTC, 2016), excessively stringent regulations that needlessly block or otherwise impede the use of personal data can disproportionately hurt the less privileged. This includes newcomers to the platform, lower income city residents, females, and those who are physically or otherwise impaired and in need of on-line medical, financial, and other services. To serve customers efficiently, producers and suppliers need to "know your customer" (KYC) well. While this is common sense, Sun et al.'s large-scale experiment shows what can happen when consumer-specific information is blocked. All participants, especially the poor and otherwise disadvantaged, incur significant welfare losses when the flow of personal data is shut off in the personalized recommendation system. Put differently, ***with proper protection, all participants are beneficiaries of the unimpeded flow of data.***

The welfare implications for e-commerce are very substantial. In China, online markets have become a new form of coordination and competition that has been helping hundreds of millions of users and tens of millions of small and micro-enterprises to follow their dreams. Online consumption in China amounted to over 10 trillion RMB in 2019, accounting for more than 25% of retail sales (about 1.5 trillion USD)[10] and has created tens of millions of jobs.[11] The costs of shutting off the free flow of personal data would be substantial. This does not justify a lack of privacy regulations. Rather, it is to say that there should be a trade-off between the benefits and risks.

(b) Digitization of trust

Akerlof (1970) has vividly demonstrated why large segments of the service economy can disappear due to the problem of information asymmetry. This is because personal information is an economic good that is subject to “the agency problems of adverse selection and moral hazard. Specifically, individuals do not know ***ex ante*** which entities have appropriate information practices (adverse selection) and ***ex post*** whether their personal information will be appropriately used (moral hazard)” (Pavlou, 2011). In the digital age, the flow and use of real-time data is a critical part of economic activity that, when properly used, can reduce the scope for opportunistic behavior, benefitting all participants of good will and good intentions. The value of data is most evident when it is used in real time. Enhancing the availability and reach of inclusive financial services, there is now the increasing ability of machine learning and artificial intelligence algorithms combined with real-time data, when and as the data are needed, and that can provide increasingly accurate and timely evaluations and recommendations to the benefit of buyers and sellers alike.

Data sharing builds trust. Trust in products and in other participants is essential for operating the online marketplace in which hundreds of millions of people make deals with each other, almost as if they were doing so face to face in the same local market. With online data sharing, customers increase their ability to rate commodities and producers. That makes producers want to build their reputations as encoded in such ratings. All participants produce and benefit from such data exchanges – in sharp contrast to offline "lemon markets" where buyers lack the information that sellers have about the goods and services they are considering for purchase (Akerlof, 1970). Just as data benefits shoppers, it allows higher-quality repeat sellers to better distinguish themselves from low-quality, "fly-by-night" sellers, shoring up their "brand," and benefitting them with stronger sales over time.

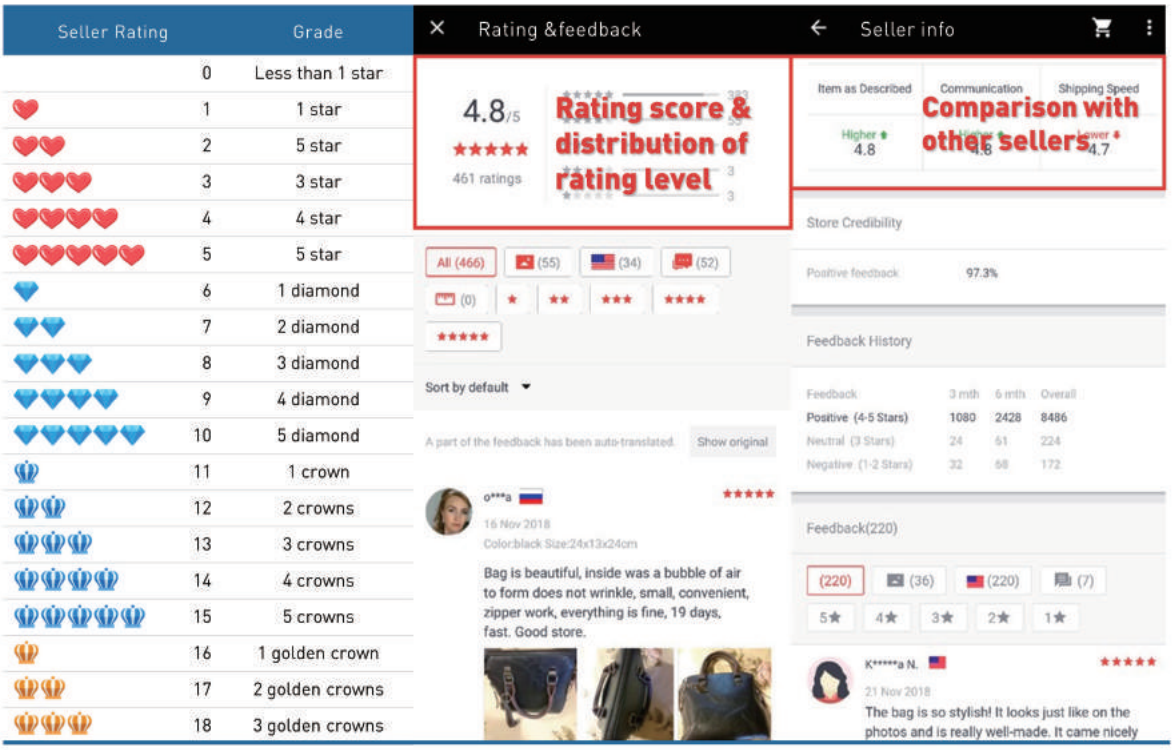
To make this happen, building trust in all items and among all participants is the key. Typically, the rating systems give buyers and sellers the right to reveal information about one other. These systems revolve around the long-term reputation of sellers, which creates an incentive mechanism for the generation of sustainable, high-quality e-retailers on a platform (Tadelis, 2002).

[10] Alibaba Announcement, May 2019, https://www.alibabagroup.com/en/news/press_pdf/p190515.pdf

[11] Reported in 2018, http://en.people.cn/n3/2019/0326/c90000-9560830.html

Taobao's 15-year experience with its rating mechanism illustrates how trust can be reliably and accurately established and sustained through the use of digital technology (Figure 20). Sellers are rated with a "star-diamond-crown" rating system. The seller can earn stars by accumulating good reviews from consumers. Five-star sellers enter into the diamond tier, and five-diamond sellers enter into the crown tier. The information used in the rating system comes from consumers—from users who are willing to share their shopping experience, as well as their post-sale experience with the use of a product. Unlike Akerlof's market for "lemons" in which there is not enough trustworthy information to separate good cars from bad cars (Akerlof, 1970), high quality sellers now, with digital marketing, can distinguish themselves and benefit from information sharing. Here Akerlof's "lemons," if not fully eliminated from every transaction, become much fewer and farther between.

Figure 20. The Rating Page View and Ranking Criteria



Seller Rating		Grade
	0	Less than 1 star
♥	1	1 star
♥♥	2	5 star
♥♥♥	3	3 star
♥♥♥♥	4	4 star
♥♥♥♥♥	5	5 star
💎	6	1 diamond
💎💎	7	2 diamond
💎💎💎	8	3 diamond
💎💎💎💎	9	4 diamond
💎💎💎💎💎	10	5 diamond
👑	11	1 crown
👑👑	12	2 crowns
👑👑👑	13	3 crowns
👑👑👑👑	14	4 crowns
👑👑👑👑👑	15	5 crowns
👑	16	1 golden crown
👑👑	17	2 golden crowns
👑👑👑	18	3 golden crowns

Rating &feedback

4.8/5 ★★★★★ 461 ratings

Rating score & distribution of rating level

All (466) (55) (34) (52) (0) ★ ★★ ★★★ ★★★★ ★★★★★

Sort by default

A part of the feedback has been auto-translated. Show original

o***a ★★★★★

16 Nov 2018

Color:black Size:24x13x24cm

Bag is beautiful, inside was a bubble of air to form does not wrinkle, small, convenient, zipper work, everything is fine, 19 days, fast. Good store.

Seller info

Item as Described	Communication	Shipping Speed
Higher 4.8	Higher 4.8	Lower 4.7

Comparison with other sellers

Store Credibility

Positive feedback 97.3%

Feedback History

Feedback	3 mth	6 mth	Overall
Positive (4-5 Stars)	1080	2428	8486
Neutral (3 Stars)	24	61	224
Negative (1-2 Stars)	32	68	172

Feedback(220)

(220) (36) (220) (7) 5★ 4★ 3★ 2★ 1★

K*****a N. ★★★★★

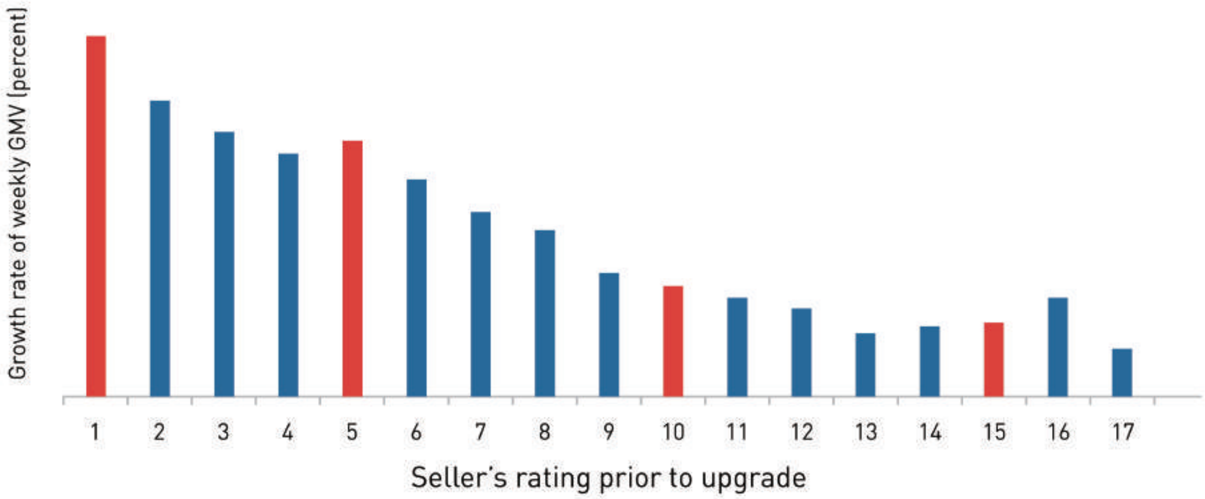
21 Nov 2018

The bag is so stylish! It looks just like on the photos and is really well-made. It came nicely

Source: ***Luohan Academy (2019)***

One way to understand the importance of data sharing to building trust is to observe what happens when a seller's rating jumps. Since the business fundamentals are continuous while the ratings are discrete, this provides an interesting window to test the value of reputation. As shown in Figure 21, what we find is that sellers usually experience a significant jump in sales in the month after a rating increase, suggesting both that reputation matters and that rating increases have informational content (there is a surprise element in the rating change). The largest jumps happen when the rating changes from zero to one heart, from five hearts to one diamond, and from five diamonds to one crown. Higher ratings are also reflected in fewer complaints and greater sales (Luohan Academy, 2019).

Figure 21. The Growth Rate of Weekly GMV after Rting Increase, 2017



Source: Luohan Academy (2019)

Note: Rating distribution is based on a random sample of sellers with positive sales amounts in 2017.

Often, we take the volume of "information flow," "capital flow," and "commodity flow" as measures of economic activity. Information flow is the sine qua non, without which capital and consumer goods cannot begin to flow from one place to another. It is an indispensable part of all economic activity. Paraphrasing Hayek, the issue of information is the issue of the economy. It connects people so that producers know how to serve customers, to build trust, and to make smarter decisions. What the digital revolution has done is to have elevated the flow of data to unprecedented levels of volume, variety, and velocity. As the use of digital technology progresses, we need to address the attendant privacy and security risks as well, a topic we turn to in the next chapter. However, one should not forget that promoting the production and exchange of information matters profoundly to the advance of human welfare, in China as well as in every place on the planet.

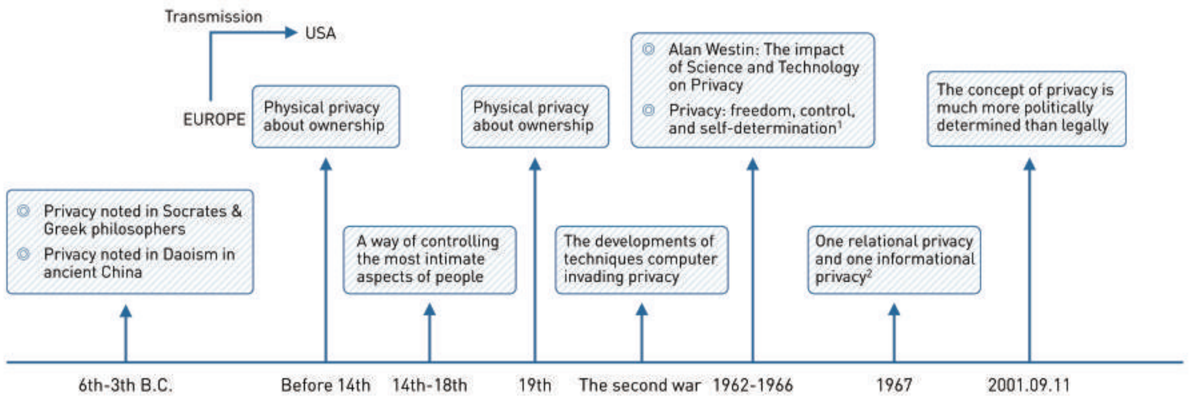
In sum, big data's larger ***volume*** and greater ***variety*** are fundamentally transforming online interactions and cooperation. They have succeeded by changing how consumers and producers connect, by increasing trust between buyer and seller, and by facilitating better and more rapid decision-making. Unlike finished commodities, data's value can only be materialized when data flows are being used. The three V's show us where the value of data comes from – large amounts of data and a rich variety of data flowing at rapid velocity drive economic activity, confirming Hayek's insight regarding the benefits of decentralized decision-making in open and competitive markets.

Chapter 4.
Privacy risks, privacy-enhancing and data safety technologies

Privacy, what the noted 19th century U.S. Supreme Court Justice Louis Brandeis called "the right to be left alone" (Warren and Brandeis,1890), underpins human dignity, liberty, initiative, and respect. The quest for privacy is a fundamental human aspiration and the right to privacy is recognized as a basic human right in many constitutions and international treaties, such as the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, and the American Convention on Human Rights. There is a broad, global consensus that individuals should not have to give up their privacy without their consent, whether implied or explicit.

Well before Brandeis' famous dictum, references to the need for personal (individual) as well as group (collective) privacy were recorded in almost all civilizations and sacred writings, from ancient Chinese and classical Greek writings (Banisar and Davies, 1999) to the Bible. Definitions of privacy vary widely across cultures, contexts, and environments, including "control," "secrecy," "intimacy," "dignity," "autonomy," "trust," and Brandeis' "right to be left alone" to name a few. All these aspects underscore the fact that privacy is among the basic and universal needs of every human being. As Volio (1981) asserts, "in one sense, all human rights are aspects of the right to privacy" (Figure 22).

Figure 22 The History of Privacy


Transmission
USA
EUROPE
Physical privacy about ownership
Physical privacy about ownership
Alan Westin: The impact of Science and Technology on Privacy
Privacy: freedom, control, and self-determination[1]
The concept of privacy is much more politically determined than legally
Privacy noted in Socrates & Greek philosophers
Privacy noted in Daoism in ancient China
A way of controlling the most intimate aspects of people
The developments of techniques computer invading privacy
One relational privacy and one informational privacy[2]
6th-3th B.C.
Before 14th
14th-18th
19th
The second war
1962-1966
1967
2001.09.11

Source: Ballard (2013); Holvast (2007) ; Dempsey (2019), summarized by Luohan Academy.

Yet, if a balance is to be struck between the benefits and costs of digitized information, privacy cannot be defined as an inalienable right. Rather, "privacy" must be viewed as the right to control and profit from one's own information (Schwartz, 2004). Consumers must be allowed to give up some of their privacy in order to enjoy many of the benefits of digitized marketing, finance, health care, education, social networking, and all the other services that have so greatly contributed to the development of China’s inclusive economy. In other words, privacy must be viewed as an exchangeable commodity that entitles its involved participants to engage in mutually beneficial trade. As another noted U.S. jurist, Richard Posner, has pointed out, too many privacy advocates have conflated what he called "seclusion" – Justice Brandeis "right to be left alone" – with "secrecy," the right to control information about oneself (Posner, 1979).

4.1 Where do privacy risks originate in the digital age?

The good news is that today digital technologies increasingly are being used to turn little data into "big data," matching buyers with sellers, allowing them all the benefits that online services have to offer. The bad news is that significant risks of privacy breach and loss of data security exist at every stage of the data flow cycle, from data collection to storage, analysis, use, and even erasure.

Consider ***data collection.*** It is widely accepted that individuals should have the right to know about and consent to data collection at this very first stage of the data flow cycle. In practice, it is a daunting challenge to protect individuals from excessive or unauthorized data collection – from every possible trick hackers and phishers may have up their sleeves.

For example, in 2018, a company in China, Ruizhi Huasheng, was found to have illegally collected 3 billion records of personal data. The company worked with local network providers for marketing, which allowed it to remotely log in to their operating systems. It then embedded data collection programs into the providers' systems. These systems collected customer data, including cookies that contained accounts and passwords, together with stored data in local servers. With these data, Ruizhi Huasheng was able to log in to numerous customers' accounts on multiple platforms. The company realized huge illegal profits by providing marketing services for many social networking platforms at a price of about RMB 0.5 per user. Even worse, some of the personal data it sold were used to engage in financial fraud.

Numerous methods can be used to collect users' data without their consent. Phishing occurs when someone collects personal data by mimicking a trustworthy entity. Phishing methods include a spam e-mail with links, a pop-up in the browser, or a phone call. The purpose of such personal data collection most often is financial theft. According to Kaspersky Lab, 12.1% of its users experienced an assault, while its anti-phishing system blocked over 100 million attempts to direct its users to scam websites in the first quarter of 2019 alone.[12]

Tracking technologies were developed as ways to find out diverse customers' real interests in order to satisfy their unique wants and needs (Hoofnagle et al., 2012), but users are often uninformed about the technologies behind online advertisements (Smith et al., 1999). Phishers and hackers can exploit these same technologies to steal from and otherwise defraud innocent victims, who are kept in the dark about the extent to which, by whom, and in what dimensions their data are being collected (McDonald and Cranor, 2010). What's more, for many of them there is no way for them to opt out of tracking.

Next consider the data storage stage. Personal data may be stored and pooled either in local servers or on an i-cloud. At least one report indicates that during 2016-2018 there were breaches involving 11 terabytes of data records.[13]One of the most notorious cases was the 2018 Facebook-Cambridge Analytica data scandal that harmed millions of Facebook users. Cambridge Analytica had illicitly harvested personal data from Facebook users without their consent, and the data was eventually used for political purposes. Its App, "This is your digital life," asked for consent from Facebook users to complete a survey for academic purposes.

[12] See https://securelist.com/spam-and-phishing-in-q1-2019/90795/.

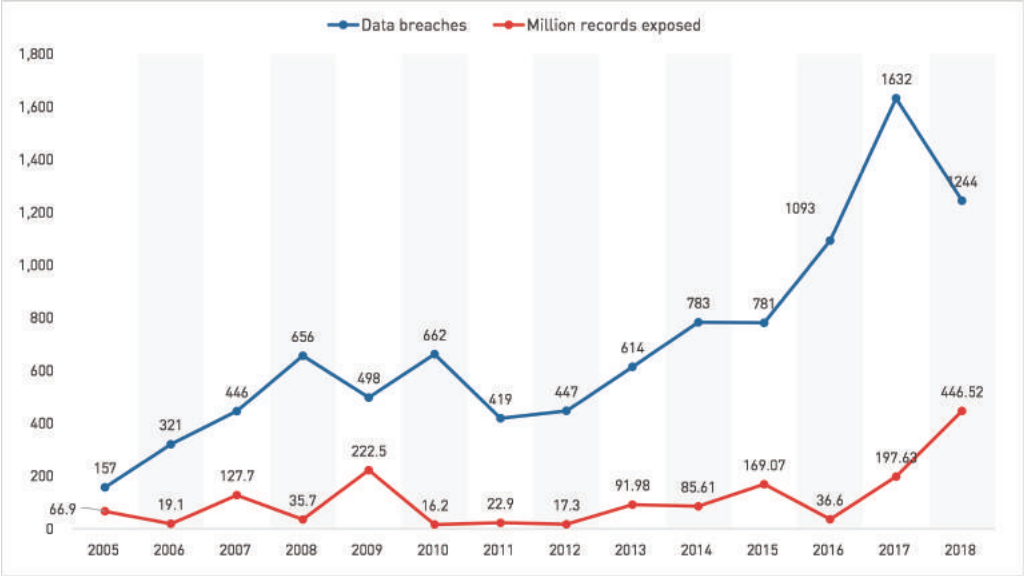
[13] See https://www.ibm.com/downloads/cas/ZGB3ERYD.

Information from all those in the consenters' social network was collected by the App and then misused. The scandal led the U.S. Federal Trade Commission to impose a USD 5 billion fine -- the largest as yet levied anywhere in the world -- along with stiff new privacy regulations on the company.[14] Both the crime and the punishment damaged Facebook's reputation.

Facebook also encountered an attack in September 2018 that exposed 50 million users' accounts, the largest breach in the company's history.[15] Worse yet, another data leakage occurred in December 2019, exposing more than 267 million Facebook users' information to an online database found in a hackers' forum. The personal data included users' Facebook IDs, full names, and phone numbers.

Facebook and its users are not the only victims. Other companies and their clients have also experienced data leakages. In the United States, the number of data breaches doubled from 2014 to 2017 (Figure 23). For example, the Verizon Data Breach Investigation Report (2015) counted more than 2,100 cases, with more than 700 million leaked records in 2014.

Figure 23. Data Breaches and Exposed Records Rose in the United States



Source: Statista

Note: Annual Number of Data Breaches and Exposed Records in the United States from 2005 to 2018 (in millions).

Finally, consider the data erasure stage. The right to be "forgotten" is an important aspect of privacy protection and online search platforms appear to have achieved significant progress in this direction. For example, user browsing history can be erased on Google Chrome, Bing, and other browsers.[16] Once the information is deleted from one device, it should be cleared everywhere it is synched. Given that the data are essential to online tracking and related to personal online identity, their erasure provides assurances to customers, who would not want their searches to be tracked. Another example is Microsoft's InPrivate Browsing feature, which allows customers to have their online behavior unrecorded by a browser (and therefore shielded from oversight) despite some inevitable inconvenience to customers.

[14] See https://www.ftc.gov/news-events/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions

[15] See https://www.nytimes.com/2018/09/28/technology/facebook-hack-data-breach.html

[16] See https://support.google.com/chrome/answer/2392709?co=GENIE.Platform%3DDesktop&hl=en

4.2 Privacy engineering and privacy-enhancing technologies

Even though today's online users are, consciously or unconsciously, trading off their personal privacy for the option to "shop and save" and otherwise enjoy the benefits of the Internet, they may not always have to do so. Of course, hackers and phishers are never going to go away, and they will always be "innovating" new ways to pick consumers' pockets. But history also suggests that new technologies can turn challenges – even the most daunting ones – into solutions. As with new drugs, new technologies can be made more sustainable by limiting the potential adverse side-effects. Importantly, there is no evidence that stringent regulations and massive fines are the only or even the best way to achieve the proper balance between the inclusive growth offered by digital technologies and the protection of personal privacy.

Because resources are limited, there can never be absolute guarantees of personal privacy. Prevention would exhaust too many resources that could be used to prevent other types of crimes, leaving aside the advancement of a seemingly infinite array of individual and societal enhancements that might otherwise be achieved.

The key to privacy protection is obviously to implement protective mechanisms and technologies that can be empowered by big data, just as the online payment system has been empowered to limit fraudulent payments. We now turn to two recent promising and complementary advances in the practice of privacy protection: ***privacy engineering ("privacy-by-design")*** and ***privacy-enhancing technology.***

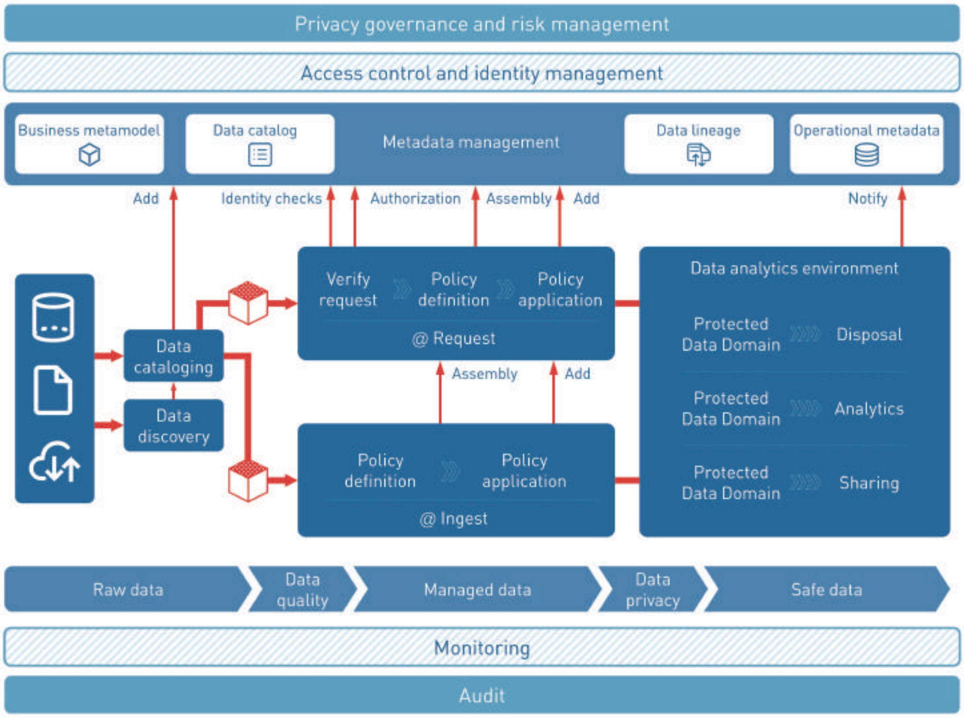
Privacy protection is becoming a core requirement for companies in the digital age. Many technology companies have embraced the "privacy-by-design" approach. Advances in privacy engineering have brought user-oriented principles into the design and use of software and service requirements. Privacy engineering encompasses two components: the design and implementation of software that allow service providers to provide protection; and user interface designs that ensure users' understanding of the privacy clauses and the extent to which privacy engineering technologies are able to protect the corresponding, sensitive information (Rubinstein and Good, 2013). Both components have been increasingly adopted in privacy protection practices.

Privacy engineering guides data recorders, processors, and software developers to translate core privacy principles into concrete design features and methodologies. Building on the work of Gürses et al. (2011), Hoepman (2014) has identified eight ways to design privacy into the software "minimize, separate, aggregate, hide, inform, control, enforce, and demonstrate." For each design strategy, an appropriate Privacy-Enhancing Technology (PET), discussed below, can also be applied, which developers can use to implement "privacy design patterns," i.e., commonly recurring structures that solve a general design problem within a particular context.

In every case, the basic strategy is to restrict the collection and processing of personal data to the minimum amount necessary. In addition, data producers are required to obtain authorization from users before collection and then to anonymize the data using pseudonyms before they are analyzed and put to use.

Today, privacy engineering is being adopted by more and more Internet services. For example, Privitar, a platform that aims to protect privacy in the lifecycle of sensitive data, has applied a set of user-centric principles to everything they do (Figure 24). They have adopted a three-stage data privacy pipeline in order to enable organizations to design data flows that automate best practices for privacy protection. Data move through the three stages of "raw," "managed," and "safe" data. Raw data, collected from a business metamodel with user authorization, is treated as high risk until the scope of any personal information it contains is well understood. Access to raw data is tightly controlled. Raw data becomes "managed" data through the processes of data cataloging, encryption, and de-identification. Applying privacy transformations to managed data creates a protected data domain containing safe data to be used by specific analysts for specific purposes. Privacy risks are minimized in this environment.

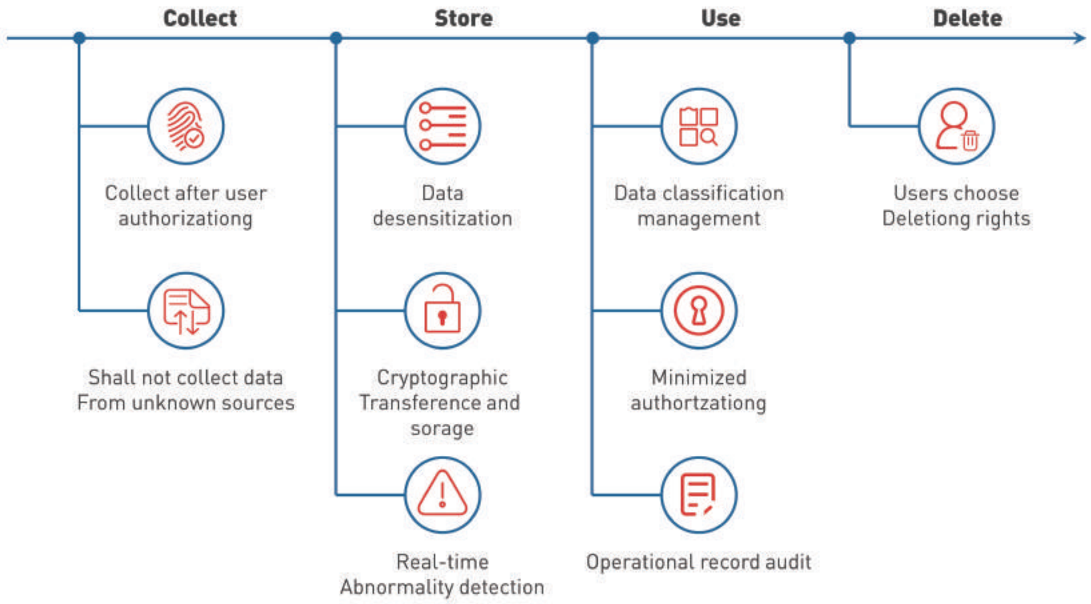
Figure 24. Privacy and Data Protection in the Data Processing Lifecycle at Privitar, a Platform that Manages the Lifecycle of Sensitive Data



Source: Privitar.com

Similarly, Ant Group has applied a set of user-centric principles to their entire data use procedure (Figure 25). At the ***data collection stage,*** user authorization must be obtained, and it must be determined that the requested data are in fact necessary. Ant prohibits the collection of data from unknown sources. In the ***storage stage,*** before any data can be used, they must be desensitized (rid of sensitive information). They must also be encrypted so that they cannot be used in the event of a data leakage. Ant Financial uses real-time, 24/7 monitoring to detect abnormal activity so as to minimize privacy risks. At the use stage, the encrypted, desensitized data can be used while under data classification management. At this stage, the engineers must balance the need for privacy protection with generation of information that will truly benefit the end users. Finally, users can always choose to exercise their deletion rights.

Figure 25. Privacy and Data Protection of the Data Processing Lifecycle at Ant Group


Collect
Store
Use
Delete
Collect after user authorizationg
Shall not collect data From unknown sources
Data desensitization
Cryptographic Transference and sorage
Real-time Abnormality detection
Data classification management
Minimized authortzationg
Operational record audit
Users choose Deletiong rights

Source: Luohan Academy

Apple follows four principles that govern the lifecycle of data processing: the minimization of personal data collection, on-device processing that limits unnecessary flows, transparency and control through authorization, and secured processing. These principles aim to reduce privacy risks in the processing of any data flows. Apple also pioneered the use of "differential privacy," by injecting noise into its data sets. For example, to figure out what emoji users like without violating their privacy, every time a user hits an emoji, another emoji is also sent to the dataset together with the used one. Without one-on-one tracking of entire user activity, the data were still sufficient to provide the information needed for useful analysis.

PETs are techniques and tools used to provide privacy protection from untrusted and potentially harmful data controllers (Gürses et al., 2011) and, as indicated, can complement privacy engineering designs. They can be classified into “hard PETs” and “soft PETs.” Hard PETs utilize complex technologies to reduce the risk of misplaced trust in a third party. One particular example is the use of a small but powerful set of cryptographic protocols to mitigate the risk of personal data disclosure. They include anonymous communication channels (that hide users' IP addresses from service providers, while allowing communication); selective disclosure credentials (that allow users to authenticate themselves and prove that they are authorized to use a system without revealing additional information); zero-knowledge proofs (that allow one party to prove to another that a statement is true, without revealing any information other than the veracity of the statement); and secure multiparty computation (that allows group computation where only the result is revealed), to mention just a few such technologies.

A concrete example is the multiple-party calculation (MPC) technology, which is widely used to achieve multiparty computation and “zero-knowledge proof” (Box 2). It allows analysts to gain insights from data without revealing the original "raw" data, which, in turn, cannot be traced back through merged data. In this way, data collaboration by several parties can be achieved without sharing the original data, making for a robust technology that can promote data flows while substantially mitigating privacy risk.

Box 2. MPC： Secure Multi-Party Computatio

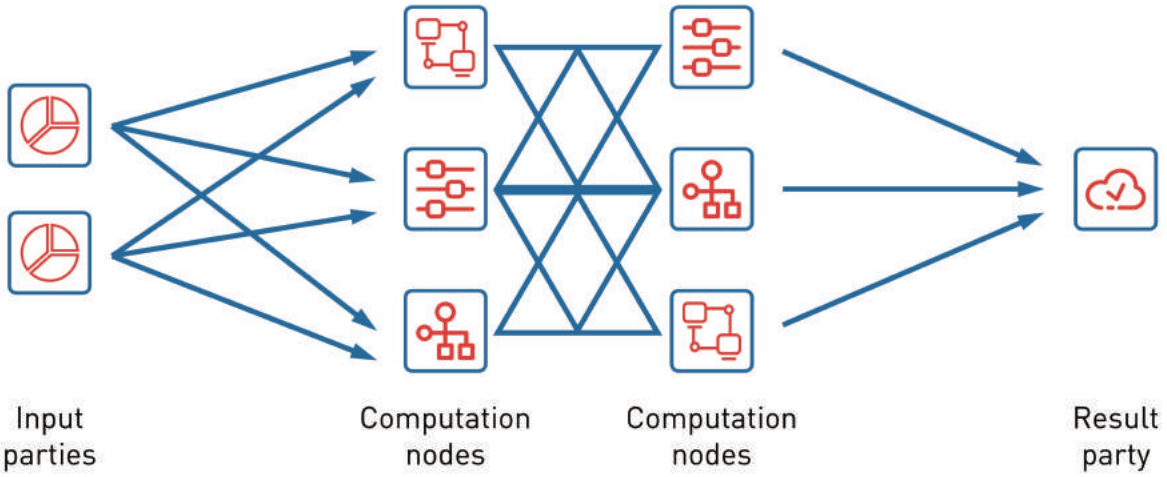
Suppose two millionaires, Adam and Bob, are interested in knowing which of them is richer. But both of them are worried about privacy and do not want to reveal their wealth. So, is there a way to tell Adam and Bob the result without violating their privacy? Secure MPC provides the answer.

This is the millionaires' problem asked in Yao (1982). In this paper, Yao proposed protocols for secure computation. With the development of big data, cloud computation, blockchain, and the awareness of privacy, the adoption of MPC has become practical in recent years. Various systems have been developed and used in real applications.

How does it work? First, data owners encrypt data onsite and upload it to the computation notes. Data analysts then build and run the queries without accessing the original data. Next, the cloud computing system processes the queries without removing the protection. The query results will be sent to authorized users in an encrypted format.

MPC is suitable for addressing the privacy concern in data use because data never existed outside the source in an unencrypted state. Various encryption methods such as homomorphic encryption, zero-knowledge proof, homomorphic commitment, and others can be used in the encryption step. The specific protocol can set standards for enhancing privacy for data from all parties.

Figure 26. The logic of a Secure Multi-Party Computation system



Source: Luohan Academy

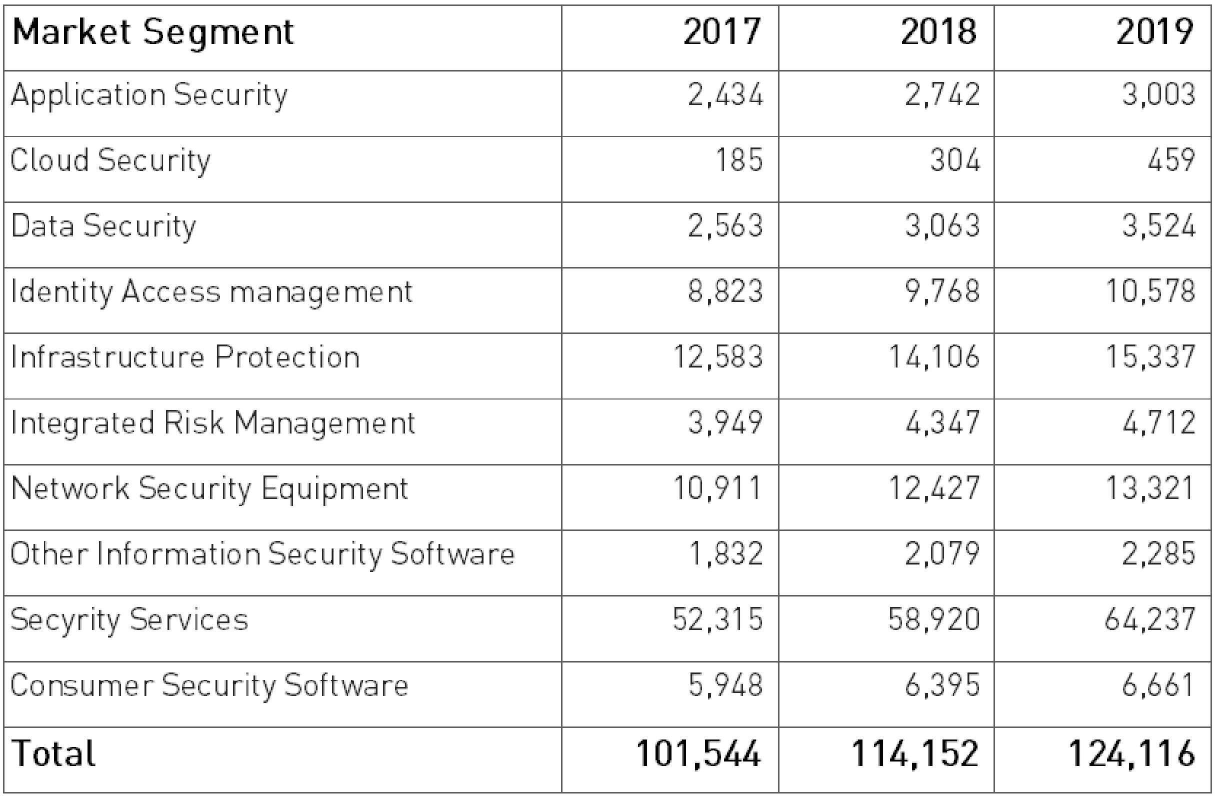
The MPC protocol can also promote the flow of data among data producers. Companies are often reluctant to cooperate with each other even when such collaboration could be very beneficial to all concerned. This happens because companies view their data as proprietary business assets and worry about data leakage during such collaborations.

The desirable features of MPC make it suitable to largely mitigate such concerns. MPC has become one of the fastest-growing technologies in recent years. Its systems are applied to many areas, including finance, education, health care, and many other businesses around the world. One example is the syndicated loan between Ant Group and dozens of commercial banks. The Secure Multi-Party Computation system allows banks and Ant Group to take full advantage of different datasets in risk management to make loan decisions without sending each other the original data. In 2019 the loans made through such collaborations grew to a total amount of RMB 2 trillion by tech companies and banks in China, benefiting finance providers, technology companies, and borrowers.

"Soft" PETs are a set of tools of data management that help users on their own to make better decisions about sharing their data with service providers while satisfying informed consent requirements, such as cookie management tools, privacy dashboards, advertising icons, and so on. The tools leave the control in the hands of the user and are based on the notion that users are able to decide for themselves how much trust they want to place in the data processor.

Hard PETs can be very costly (Figure 27). Complex distributed computing systems and encryption algorithms require significant computing power. For example, Ant's Secure Multi-Party Communication (SMC) system involves a high standard of communication efficiency and well-designed algorithms. In the computation process, there are many feedback loops among platform engineers who must go through multiple steps to arrive at a relatively simple outcome. The systems need to be designed and customized to every application separately. Setting up such a system can be very expensive and time-consuming.

Figure 27. Global Privacy and Data Security Spending Has Risen Rapidly



Market Segment	2017	2018	2019
Application Security	2,434	2,742	3,003
Cloud Security	185	304	459
Data Security	2,563	3,063	3,524
Identity Access management	8,823	9,768	10,578
Infrastructure Protection	12,583	14,106	15,337
Integrated Risk Management	3,949	4,347	4,712
Network Security Equipment	10,911	12,427	13,321
Other Information Security Software	1,832	2,079	2,285
Secyrity Services	52,315	58,920	64,237
Consumer Security Software	5,948	6,395	6,661
Total	**101,544**	**114,152**	**124,116**

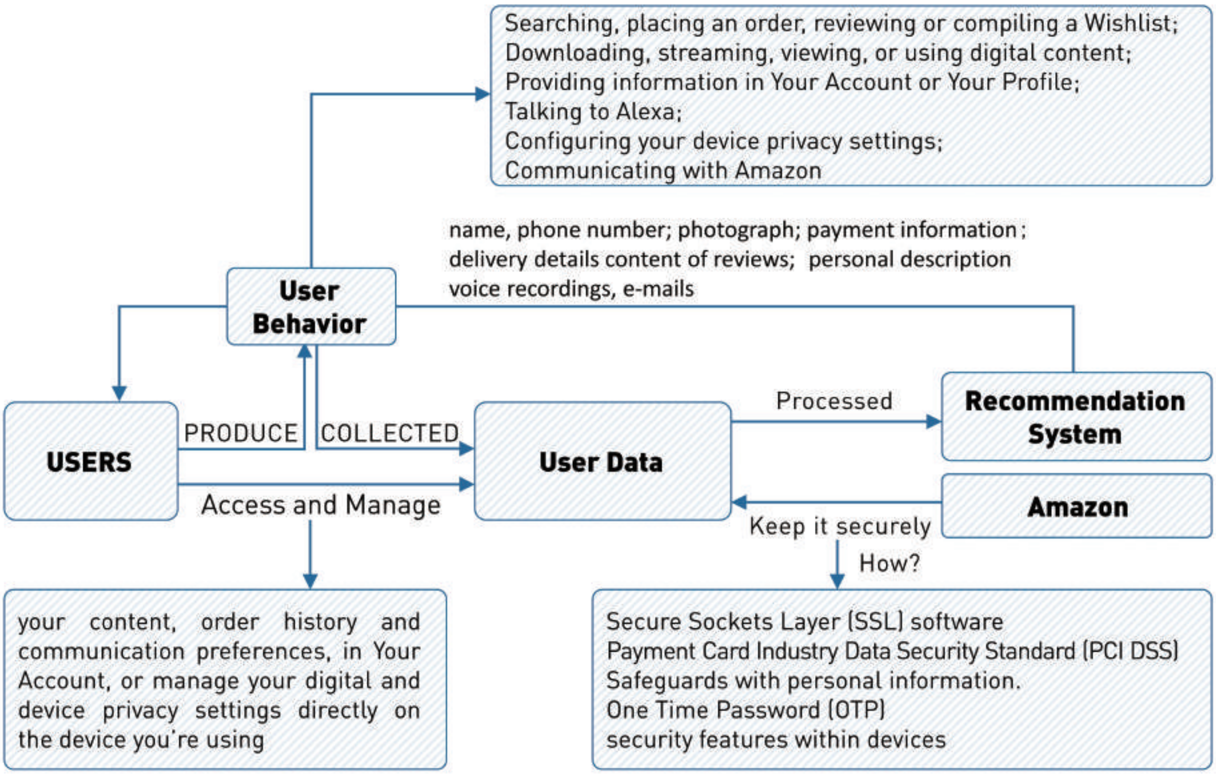
Source: Gartner (2018)

Note: Worldwide spending on information security products and services, 2017-2019, is in millions of US dollars. Risk management and privacy concerns within digital transformation initiatives will drive additional security-service spending through 2020 for more than 40% of organizations. Services (subscription and managed) will represent at least 50% of security software delivery by 2020.

At the moment, use of soft PETs is a more common business practice than that of the costlier hard PETs, which requires familiarity and expertise with the relevant cryptographic protocols. They also face significant tradeoffs with the generation of data needed for commercial applications. For now, the expensive hard PETs remain out of reach for most SMEs and startups. Soft PETs are much less costly and are both more "privacy-friendly" and "business-friendly" because they tend to enhance a firm's reputation for being trustworthy while imposing fewer restrictions on data collection and analysis. Yet advanced, highly sophisticated “hard” PETs are evolving rapidly. They are the "wave of the future." In the meantime, more and more competitors are combining privacy-by-design with PETs, soft with hard, in their digitized business designs.

Figure 28 illustrates how Amazon uses and protects personal data. Privacy protection has become a key element of competition in the digital economy as consumers enjoy more and more options for securing their personal information.

Figure 28. How Amazon Protects Privacy


Searching, placing an order, reviewing or compiling a Wishlist;
Downloading, streaming, viewing, or using digital content;
Providing information in Your Account or Your Profile;
Talking to Alexa;
Configuring your device privacy settings;
Communicating with Amazon
name, phone number; photograph; payment information;
delivery details content of reviews; personal description
voice recordings, e-mails
User Behavior
USERS
PRODUCE
COLLECTED
User Data
Processed
Recommendation System
Amazon
Access and Manage
Keep it securely
How?
your content, order history and communication preferences, in Your Account, or manage your digital and device privacy settings directly on the device you're using
Secure Sockets Layer (SSL) software
Payment Card Industry Data Security Standard (PCI DSS)
Safeguards with personal information.
One Time Password (OTP)
security features within devices

Source: Amazon.de.

Note: Amazon's illustration of how it uses and protects personal data.

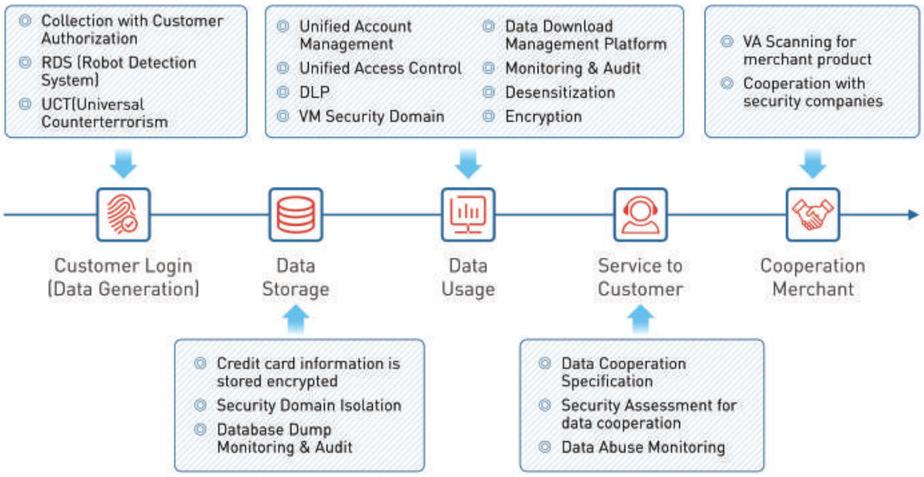
4.3 Data Security

Digital service providers must keep their data secure. This requires a strong measure of self-governance together with the application of the latest advances in security technologies. Industry "best practices" for monitoring and risk management, the basis for effective self-governance, provide these firms a guide with which they can keep their data secure throughout the big data lifecycle. Industry self-governance can also include independent certification agencies, industry codes of conduct, stakeholder participation on corporate boards, and so on.[17] Many designs and technologies serve up-front privacy and downstream security protection at one and the same time.

Multiple techniques and tools are being deployed to enhance the security of data (Figure 29). They include access control platforms, data classification, desensitization tools, audit platforms, encryption tools, and so forth. These techniques and tools can be applied at every stage of the data lifecycle. Data security techniques and tools are being constantly upgraded along with the broader advances in digital technologies. One of the recent examples is cloud computing that facilitates data analysis without storing the process offline.

[17] For an extensive discussion of such options and their potential effectiveness, please see OECD (2015) on industry self regulation.

Figure 29. Risk Management Over the Data Life Cycle


Collection with Customer Authorization
RDS (Robot Detection System)
UCT(Universal Counterterrorism
Unified Account Management
Unified Access Control
DLP
VM Security Domain
Data Download Management Platform
Monitoring & Audit
Desensitization
Encryption
VA Scanning for merchant product
Cooperation with security companies
Customer Login (Data Generation)
Data Storage
Data Usage
Service to Customer
Cooperation Merchant
Credit card information is stored encrypted
Security Domain Isolation
Database Dump Monitoring & Audit
Data Cooperation Specification
Security Assessment for data cooperation
Data Abuse Monitoring

Source: Luohan Academy

Figure 30 illustrates the risk management process used by the Alibaba Group. Constructing such a self-governance framework for data security is essential. It enables firms to avoid unnecessary risks and to respond quickly to unexpected incidents. It starts with an emphasized consensus from the management and entails organizational support. For example, at least 28,000 data protection officers (DPO) were hired by firms in the EU and the US after the GDPR came into force in 2018.[18] One company has established a four-layer organization to protect data security, from strategy to management, internal controls, and execution. Its data security team accounts for 2% of the total number of employees. The company provides an Information Security Awareness Program to all employees in order to enhance data protection.

Figure 30. Data Security Management Tools

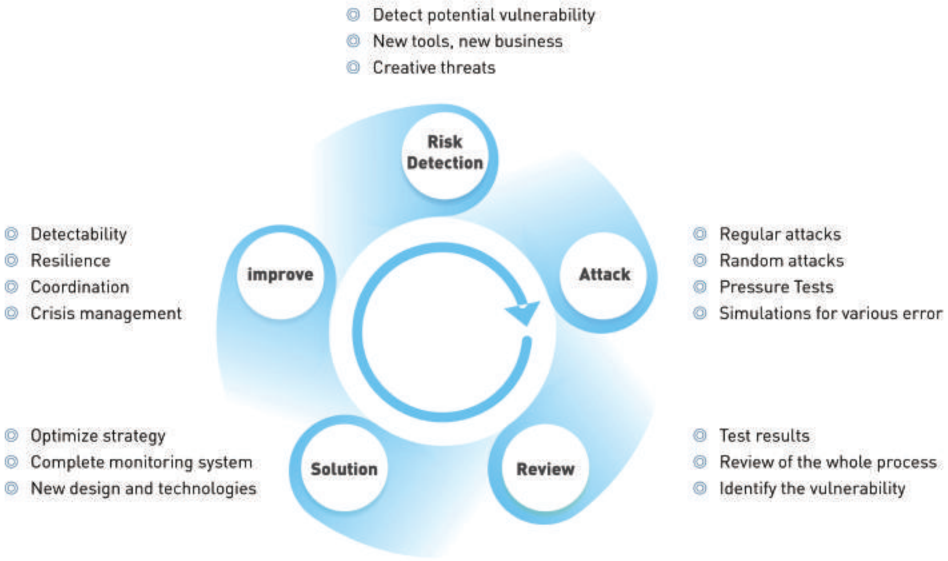
Network	Firewall Network access authorization tool Honeypot network Drop traffic detection (black hole) IDS WAF
Terminal Device	File Encryption System DLP (Data Leakage Prevention) Anti-virus Software
Data Process Platform	Data Download Management Platform VM Security Domain
Access Control	Access control & Authentication Platform
Operation Monitoring	Security related Risk Operation platform
Encryption	Encryption Equipment Key Certificate Management Platform

Source: Luohan Academy

[18] See https://iapp.org/news/a/study-at-least-28000-dpos-needed-to-meet-gdpr-requirements/.

Cybersecurity and tech companies like Ant Group use "war games" with "opposing forces" to test and improve their systems, making sure they can quickly respond to scenarios where data and privacy breaches threaten to severely damage the reputations of their retailers. Ant Group has an "opposing force" unit, called the ***Site Reliability Engineer*** (SRE, Figure 31), whose task is to constantly look for and exploit vulnerabilities, regularly "attacking" the data and privacy management system. The targets of these tests include data security, algorithm performance, cloud computing, and intermediate development software. The “war games” are carried out at different levels. There are regular but random simulations, in which a business unit encounters assaults by the SRE team. Each year, a special month is devoted to data security, when any unit within the company can face random attacks from the "opposing force." Such exercises are not limited to cyber-attacks on technological infrastructure. Since 2017, SRE has added physical threats to its exercises, such as simulating natural disasters and assessing their impacts on the viability of the platform.

Figure 31: Ant Group Uses an Opposing Force Unit to Keep Improving the System



Source: Luohan Academy

Overall, both privacy protection and data security call for a comprehensive framework for integrating technology and user-oriented design for self-governance. It is increasingly becoming a core competence necessary for harnessing the benefits of the digital revolution. Data privacy and security issues can be significantly mitigated through such mechanism design and technological solutions. Just as with the evolution of health and safety in the food industry, when the right technology is in place, a greater amount of data does not have to necessarily imply greater privacy risks. It is quite the contrary. With growing concerns over data privacy and security, more and better technologies and mechanisms will become available over time and grow into a core competitive strength of many institutions in the digital age.[19] We, the in-house co-authors of this report, expect the costs of such technologies to go down rapidly, promoting privacy-protection-as-a-service (PPaaS) and data-security-as-a-service (DSaaS) that can benefit millions of small companies. Importantly, and as noted in Chapter 6, these are profit opportunities for innovative tech companies and online retailers who wish to capture market share from those who might not be so inclined.

[19] Regulators, such as the Bank for International Settlements, started to emphasis the important role of tech in regulations. See Cœur é (2020) for a great discussion about regtech and suptech.

Chapter 5.

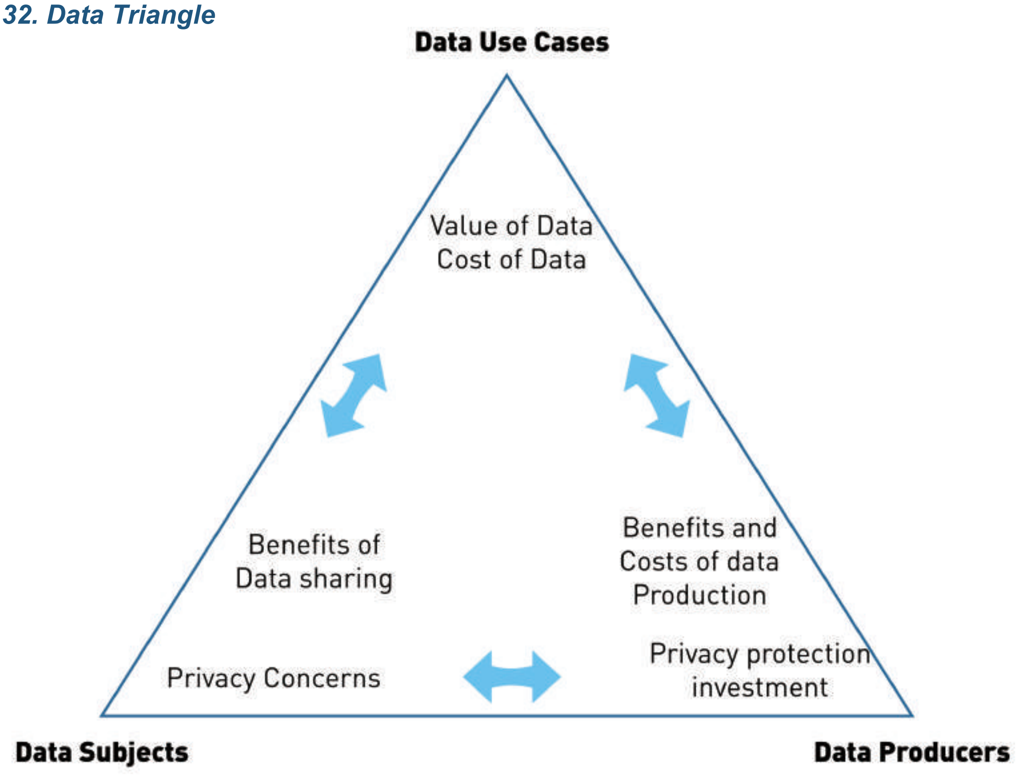
An integrated framework for understanding the data calculus

Having discussed the benefits of personal data sharing and the privacy risks, and privacy engineering and technologies that are being developed to address them, it is time to think critically about how we might better understand the usefulness of what we call the data calculus.

This requires an integrated framework. Misperceptions of big data are common, partly because of narrowly focused concerns about privacy and security, and partly because of a failure to understand the essence of the meaning of big data and thus to fully comprehend the actual tradeoffs.

We offer here a simple data triangle that we think will enhance the reader's understanding (Figure 32). Note that that it consists of data subjects, data producers, and data use cases. The term, "data subjects" refers to the parties (whether for commercial or other applications) from which the data are produced. "Data producers" refers to parties that collect, process, store, or distribute data. "Data use case" refers to real-life scenarios in which data are produced, processed, and utilized to facilitate economic or social activities.

Figure 32. Data Triangle



Source: Luohan Academy

Figure 32 shows how the essence of data can be captured in three pairs of relations.

First, data producers and subjects are intertwined through data production, use, benefit distribution, and risk management.

Information is not only a key determinant of production and trade. Encoded information also has the fundamental property of non-rivalry (Arrow, 1969), as we have noted. Unlike a physical commodity such as oil, information can be produced and used an unlimited number of times without depleting the original material. Unlike oil, data is more akin to fire which can be passed on. Its value grows as it is more widely shared. As Romer (1990, 2018) explains, information is a special type of production input in promoting human progress and this special property of non-rivalry is not always sufficiently appreciated. Due to its non-rivalry feature, data can potentially be produced or used an unlimited number of times, simultaneously or in sequence, without depleting the value of the data.

By production, we mean observing, recording, and processing the data. While data subjects (individuals) can report data, in most cases, especially with the advance of digital technology, data usually are observed or inferred by third parties. For example, the fact that someone is attending a meeting is a piece of data observed by the eyes and ears of fellow participants. The data is not only produced by the data subject, it is also produced by others.

Each participant is likely to produce a different view or version of the data that can vary in minor details or even in essence. The process of consumer search and purchase offers yet another example. Data subjects usually have no intention to produce the data; nor do they incur any direct costs in its production.

There is a large gap between simple data creation – the "data subjects" – and the potentially insightful information that can be extracted from them. For example, data cleaning is an important first step in any machine learning application. The “cleaners” have to clean the data because they know what they need to learn about. This means that the value of a data set is in the eye of the beholder, for example, the sorts of things she can learn about (Blackwell, 1953).

More importantly, it means that processing data into valuable information is costly. Processing requires the use of “scientific methods, processes, algorithms and systems to extract knowledge and insights from” the data subjects. “Information” is not raw data. It is not the province of the bean counter. Rather, it requires the skills of a highly skilled data engineer, who “uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data. Data science is related to data mining, machine learning and big data.”[20]

Data are not pieces of in-shape commodities that data engineers carry around in their pockets. Rather, in the digital world they are produced each time the activities of data subjects are observed and subsequently processed. While data subjects, being aware of their own activities, have their specific concepts of data, so do other observers. Depending on how data is observed and processed, there can be very different versions of the same raw set of data.

The non-rivalry nature of data production explains why it would be inefficient to assign sole ownership to individual data subjects. Data subjects do not usually take into account the positive externalities data can generate for all other data producers and users. There can potentially be an unlimited number of

[20] https://en.wikipedia.org/wiki/Data_science

"owners" of data, without necessarily infringing on the rights of any of them. Assigning ownership to just one or a few of them would mean massive under-provision of useful information.

We have given many examples of the value of data enjoyed by individual data subjects without assigning to them sole ownership. That is not the end of the story. The non-rivalry nature of data use also implies that data producers can use data in such a way that hurts the privacy of the data subjects or neglects the security of their data, what we call here the problem of "non-separability". This is where privacy and security issues arise and, of course, there is no lack of examples of data privacy and security breaches in real life.

Therefore, data subjects and producers are inevitably intertwined in the production and use of data. A meaningful and sustainable solution is not to lock up data through sole ownership, but to provide incentives to all participants who can contribute to the production and sharing of useful data, while protecting their privacy and security rights.

Second, the value of data is an indispensable part of the life cycle of big data

It is tempting to think of data solely as a commodity bought and sold at a certain price at a discrete time and place. In reality, the value of data is realized as an indispensable by-product of ongoing economic activities that, in principle, benefit everyone involved. Take the example of mobile taxi-hailing. The digitization of connectivity has created a new market in which riders and transportation providers can break the traditional boundaries of physical location to find each other in a timely and safe fashion. To achieve this, all participants must release information about their identities and ongoing real-time locations. Such data make the service possible, and all participants – including the riders, the vehicle providers, and the digital platforms (online services) that have created this market – are beneficiaries of the exchange of information. The value of information is realized in real-time even as the service is being provided. It is reflected in the increased mobility provided to riders, the earnings provided to drivers, and the revenues provided to platform operators that have put the application together.

The bike-sharing business offers yet another example. By releasing digitized information of their identity, bicycle riders can now rent bicycles on the spot. The exchange of data is a necessary condition for this to happen. Bicycle riders and providers both benefit from this process as do the platforms that make it possible.

Third, while data sharing raises privacy risks, with proper mechanism design and competent technology (see Chapter 4), the tradeoff between risks and benefits can be made manageable.

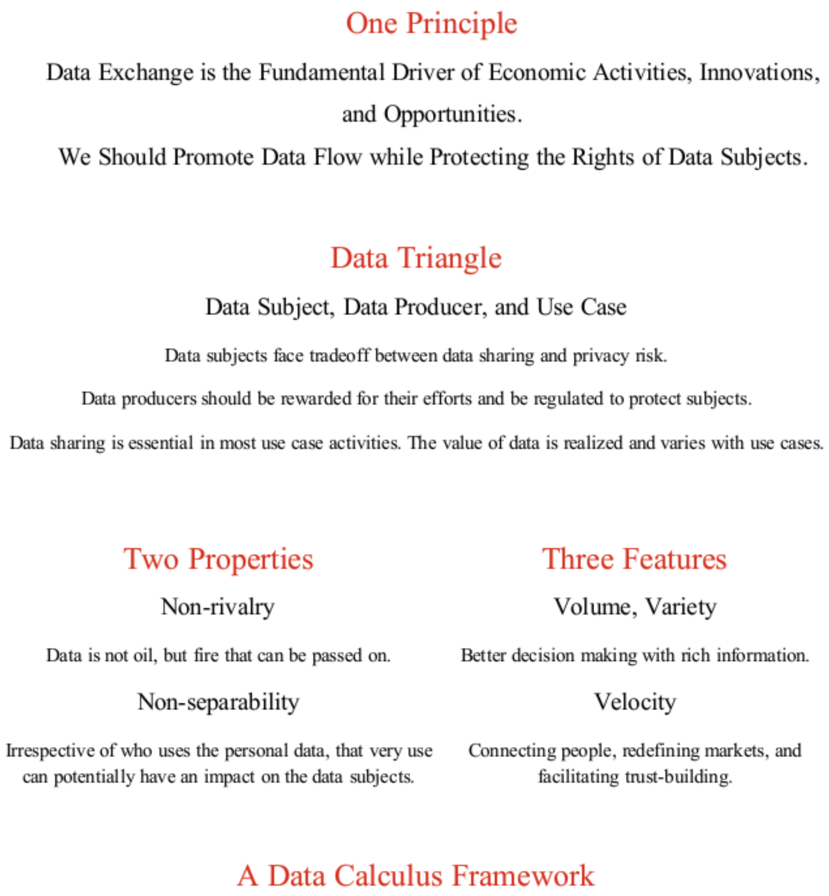
Analogously, travel by airplane and automobile raises concerns about mortality risks and those risks will never go away. But with competent flight and automotive technology, together with government and industry self-regulation, few people nowadays will avoid travel by either mode because of a tradeoff they have made between mobility and risk. The same is true for travel by elevator, where technologies have all but eliminated the risk of death. Similarly, with sufficient regulatory requirements and competent processing technology, there is little relation between the amount of food one consumes and the risk one is going to get poisoned. Digital privacy protection may never reach the protection afforded for elevator travel, but that should never stop the search for perfection.

Historically, breakthrough technological innovation has seldom occurred without raising new challenges, just like the concerns about airline, passenger, and even elevator safety. For the benefits to be sustainable, technology has nearly always found solutions that reduce the attendant risks to acceptable levels.

Information processing and sharing has always been a fundamental part of human progress. It is only since the volume of information sharing has risen to present-day heights, with the unprecedented coordination made possible by digital platforms, that the downside risks have become so concerning. But as in most industries, with the proper design of mechanisms and with competent technologies, we can hope to get the benefits of online commerce while reducing privacy risks to more manageable levels. The real challenge then is not to shut down data sharing, but to achieve this goal in an effective, efficient, and sustainable way.

We should not deem privacy risk to be an insurmountable obstacle to the further evolution of personal exchange within the digital marketplace, with all its promise for inclusive prosperity. The real choice is not between two polar extremes. Rather it is to use a data analytical framework (Figure 33) to strike the right middle ground – the right balance between the benefits of promoting data flows and the costs of privacy and security protection, even as the balance is changing with improving mechanism designs and the corresponding technologies undergirding them.

Figure 33: An Illustration of the Data Analytical Framework



One Principle

Data Exchange is the Fundamental Driver of Economic Activities, Innovations, and Opportunities.

We Should Promote Data Flow while Protecting the Rights of Data Subjects.

Data Triangle

Data Subject, Data Producer, and Use Case

Data subjects face tradeoff between data sharing and privacy risk.

Data producers should be rewarded for their efforts and be regulated to protect subjects.

Data sharing is essential in most use case activities. The value of data is realized and varies with use cases.

Two Properties

Non-rivalry

Data is not oil, but fire that can be passed on.

Non-separability

Irrespective of who uses the personal data, that very use can potentially have an impact on the data subjects.

Three Features

Volume, Variety

Better decision making with rich information.

Velocity

Connecting people, redefining markets, and facilitating trust-building.

A Data Calculus Framework

Source: Luohan Academy

To sum up, our data analytical framework is consistent with ongoing digital developments and the growing need for more and better digitized services based on the rapidly growing exchange of personal information. The growth of the digital economy is producing greater and greater flows of digitized data, with unprecedented levels of participation, collaboration, and new, much more inclusive opportunities for businesses and their customers – for buyers and sellers, rich and poor, everywhere in the world. There is also a growing need to promote privacy and security engineering in order to better establish and maintain trust in the system and to make the digital economy work for everyone.

But with the ongoing advance of technologies such as those discussed in Chapter 4, it is becoming increasingly possible to promote data sharing without incurring insurmountable privacy and security risks. Data consumption and privacy risk do not have to be treated as "either/or." ***The key challenge in the digital age is not to achieve an absolute goal of one or the other. Rather, it is to find a common, middle ground that reaps the benefits of the digital revolution while minimizing privacy and security risks.*** This is the subject of the next chapter.

Chapter 6.

Data Governance Based on the Data Calculus

This chapter examines government and industry self-regulation of data privacy and explores some of the most recent theoretical and empirical evidence regarding the impact of "big data" on competition, innovation, and price discrimination. We argue that, in each case, the economist's tool kit – cost-benefit analysis – should serve as a guide if we are to maximize the potential for inclusive economic development.

6.1 Data privacy

6.1.1 Evolving principles of privacy protection

The United States Federal Trade Commission (FTC), the lead agency for privacy protection in that country, urges online companies to ***"draw lessons ... from the research in order to help them maximize the benefits of big data while mitigating risks"*** (FTC, 2016). That is our objective in this chapter. It is neither to maximize benefit regardless of risk, nor to minimize risk without regard to benefit. Rather, it is to help companies and policy makers within diverse cultures and with diverse tradeoffs to find the right balance – to find a middle ground that might appeal to all interested parties.

The digital information revolution has led to evolving principles for developing such a market, though specifics vary with different cultures and different preferences for the balance between information dissemination and privacy protection. As noted in the introductory chapter, modern privacy protection builds on the Fair Information Practices (FIPs), which originated from the 1973 report issued by the Advisory Committee on Automated Personal Data Systems for the Secretary of the U.S. Department of Health, Education, and Welfare (HEW). The HEW report called on Congress to adopt a "Code of Fair Information Practices" based on five principles: transparency, use limitation, access and correction, data quality, and security.[21] Based on these principles, the U.S. Privacy Act of 1974 was passed (Hartzog, 2016).

The Federal Trade Commission, America's chief enforcer of privacy rights, characterizes these principles as "(1) Notice/Awareness; (2) Choice/Consent; (3) Access/Participation; (4) Integrity/Security; and (5) Enforcement/Redress" (United States Federal Trade Commission, 1998). Whatever the specific characterization, the FIPs provide a common language about data and privacy protection. They can be used as a measure of compliance and a means of enforcement. They reflect a wide consensus about the need for broad standards to both protect individual privacy and facilitate free information flows in an increasingly data-driven economy. The FIPs started to gain international influence after they were officially adopted in 1980 by the OECD in its Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (OECD, 1981) and in 1981 by the Council of Europe in the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data.[22]

[21] The five principles include:

(1) There must be no personal data record keeping systems whose very existence is secret.

(2) There must be a way for an individual to find out what information about him is in a record and how it is used.

(3) There must be a way for an individual to prevent information about him that was obtained for one purpose from being used or made available for other purposes without his consent.

(4) There must be a way for an individual to correct or amend a record of identifiable information about him.

(5) Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuse of the data.

[22] Europe Consulting Association, Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, ETS No. 108 (1981).

Both of these organizations clearly defined personal information as data that requires protection at every stage, from collection to storage and dissemination. Their work has had a profound effect on the enactment of laws around the world, including the influential EU Data Protection Directive Principles,[23] the above-mentioned U.S. Federal Trade Commission (FTC) Privacy Principles,[24] and recent legislation of comprehensive privacy laws, such as the General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA).

The OECD Guidelines aim to "harmonize privacy legislation and while upholding such human rights, ... at the same time prevent interruptions in international flows of data." The Guidelines' emphasis on both the protection of privacy and on the need for smooth data flows is consistent with the logic of a data calculus. In Europe, the Data Protection Directive was motivated partly by an ambitious program by a group of European countries to create a "common market" and "economic and monetary union" contemplated by the Treaty of Rome (Cate, 2006).

The focus of FIPs has been evolving. While the early versions of FIPs were aimed at protecting individuals from unfair or deceptive use of information, the later sets of FIPs, particularly those adopted since the 1980 OECD Guidelines, have been targeting stronger consumer control of their personal information. Recent privacy protection legislation, including the GDPR and CCPA, strengthen consumer control. The GDPR grants data subjects eight fundamental rights[25] when it comes to the processing of their personal data. The CCPA was drafted in the spirit of five principles[26] regarding the rights of consumers, four of which focus on enhancing consumer control over how their information is used and accessed.

While in the process of ongoing evolution, the most important goal of FIP-based governance of data is not to lock up data or restrict ownership within a static framework, but rather to set dynamic, ever-improving principles to promote increasingly secure, privacy-protected data flows over time. According to U.S. International Trade Commission estimates, global digital trade, including data-processing and other data-based services, led to a significant increase in the nation’s GDP, by improving productivity and lowering the costs of trade (USITC, 2014). Overall, data flows are estimated to have boosted world GDP by about 10 percentage points over the past decade. Thus, for example, ***the GDPR lists free movement of personal data within the European Union as an important goal next to the protection of personal data.***

It is important to understand the nature of the tradeoffs. As we explained in Chapter 4, if a balance is to be struck between the benefits and costs of digitized information, privacy cannot be defined as an inalienable right.

[23] In 1990 the Commission of the then-European Community published a draft Council Directive on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data.

[24] Beginning in the mid-1990s, the Federal Trade Commission and states attorneys general encouraged U.S. operators of commercial websites to adopt and publish online privacy policies. Adoption of such policies was voluntary; compliance with them was not. The Commission interprets Section five of the Federal Trade Commission Act, which empowers the FTC to prosecute “unfair and deceptive” trade practices, to include violations of posted privacy policies.

[25] The rights are: right to be informed, right of access, right to rectification, right to erasure/to be forgotten, right to restrict processing, right to data portability, right to object and rights in relation to automated decision making and profiling.

[26] The five basic principles are: (1) Right to know about the collection of information; (2) Right to know about the use of information; (3) Right to say no to the use or sale of information;(4) Right to access information and to request removal; (5) Right to equal service and price.

Rather "privacy" must be viewed as the right to control ***and profit*** from one's own information. There are "opportunity costs" imposed on consumers – the benefits that people have to give up when pursuing a specific objective, as when the pursuit of personal privacy might require foregoing the benefits of finding helpful information on marketers of a given good or service. To quote Movius and Krup (2009): "There are clear benefits to protecting privacy, but there are related costs as well. Privacy can impose economic and social costs; while privacy may protect some individuals, it may result in costs by preventing others from making fully informed decisions (Fromholz, 2000). As countries attempt to secure privacy protection for their citizens, they move along a privacy continuum [where] the cost of providing privacy protection can be described as a relinquishment of economic efficiency and security."

It is also important to understand the role of "trust" as the way towards the middle ground. As noted by Peter Winn, chief privacy officer and Director of the U.S. Department of Justice's Office of Privacy and Civil Liberties, and put into words in (Layton 2019), "[T]rust is fundamental to the efficacy of any institution, whether a firm, a country, or the DOJ itself ... [People can be blinded] by a false choice that online privacy governance must either be a leviathan state (social control with an absolute sovereign) or a free-market system based purely on private property rights."

Trust, we submit, is, indeed, ***the*** way towards the "middle ground" on which regulation and self-governance can complement one another, to the benefit of everyone concerned.

6.1.2 Challenges to privacy protection

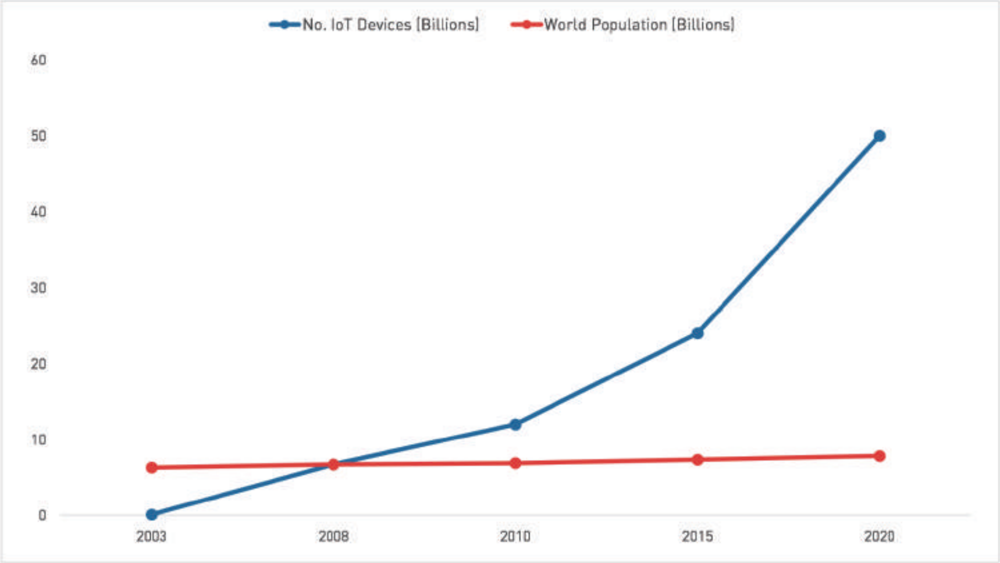
To be sure, implementation of the FIPs has struggled to strike just the right balance between privacy protection and the free flow of information. Initially, as the FIPs were introduced into national laws, they were often reduced to mere procedures such as the "notice-and-consent" approach for obtaining informed consumer consents.[27] Transparency was reduced to a notice, which required that data subjects must be made aware of what and how their personal information was being used. “Use limitation” was reduced to an overly broad and effectively meaningless “consent” that data collected for one purpose could not be used for another without the user’s permission. Indeed, the notice-and-consent approach has become increasingly unworkable as the foundation for modern-day privacy protection policies.

The explosive growth of data flows has burdened businesses with challenging legal obligations, even as users have had to endure an onslaught of unfathomable notices and often-limited choices. "Notice," the "fundamental core principle" in the FTC's Privacy Principles, makes little sense in situations involving a massive amount of data – situations in which obtaining and giving consent have become increasingly cumbersome, given the growing complexity and number of decisions that must be made at each and every moment in time. As observed in Holdren and Lander (2014), "Only in some fantasy world do users actually read these notices and understand their implications before clicking on consent." For example, given the explosive growth in the Internet of Things (IoT) – "the millions of Internet-connected devices that are in the market " (FTC, 2016), the burden of reading notices and giving consent has become far too great for consumers to be able to control their personal data.

[27]. Cate (2006).

And if the cost of obtaining consent becomes too great to make the proposed use of data economically and practically feasible, there will be nothing to which the consumer can consent, leaving nothing but the "privacy paradox" we started with in Chapter 2. Alternative approaches need to be designed to accommodate the still quickly increasing data flows.

Figure 34. Explosive IoTs, Explosive Notice-and-Consents



Source: Cisco

Second, big data is not only "big" – it has a brain. It can be made "smart." It can be used to provide an abundance of information that is beyond imagination. But it can also be used to generate additional, heretofore unimagined threats to personal privacy. Consider, for example, the current notice-and-consent approach. From the history of transaction patterns, a teenager's pregnancy can be predicted from her vitamin purchases. One-night stands can be inferred from discerned patterns in data on pick-ups and drop-offs. Such [28]misuse of data cannot be prevented by the simple procedures of notice and consent. A [29]"good" firm may be conscious of the potential threats and not abuse the data in any such way. A "bad" firm may be tempted to take advantage of the data; it can easily mislead customers or users while "complying" with requirements. In order to protect data subjects from the bad firms, the good firms may suffer significant compliance costs, an example of Akerlof's problem of opportunistic behavior and moral hazard.

Third, the expansive network of big data can create significant negative externalities, giving rise to potential problems occasioned by a few players who "hold up" the rest by threatening to withdraw their consent over the sharing of critical data (Landau, 2015). And, as just mentioned, the willingness of a few individuals to disclose information about themselves may reveal the correlated traits of others (Hirshleifer, 1971).

[28] How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did, 16 February 2012, https:// www.forbes.com/ sites/kashmirhill/2012/02/16/ how-target-figured-out-a-teen-girl-was-pregnantbefore-her-father-did/#6723f52b6668

[29] UBER’S DELETED “RIDES OF GLORY” BLOG POST, Dec 2014,http://www.whosdrivingyou.org/ blog/ubers-deleted-rides-of-glory-blog-post

The choice of the consenting minority, the minority who may have the most to lose from the revelation of such information, could prevail. Consider Target's pregnancy prediction score[30] as an example. If some women choose to share their information about their pregnancy status, data analysts can then use their distinctive shopping traits to infer whether other women might be pregnant, even without the latter's explicit consent to make the knowledge public.

Such often-unpredictable results of the big data calculus can render many other rights in the FIP regulations much less meaningful. For instance, given the distributed nature of data storage, it is practically impossible for a data producer to locate "all the data about an individual," not to mention to delete them or grant access on any specified schedule in response to a data subject's request. Such an obligation could add no more than compliance costs with only an illusion of control.

All of this is to underscore the importance of ongoing efforts to develop increasingly effective technologies to mitigate such challenges. As shown in Chapter 4, these efforts are starting to show promising results. A middle ground can be found and building trust among all interested parties is the way to get there. We do not have to wring our hands in despair.

6.2 Market competition of data-driven businesses

The big data revolution is reshaping the landscape of the economy. It is important to understand its potential anti- (and pro-) competitive effects. In theory, as data have become important and unique inputs to economic activities, they can be used more easily to price discriminate and to engage in other practices that harm consumers, block innovation, and foster the generation of unfair monopoly power to keep out other, more efficient competitors. Given that data-driven businesses play an important role in economic growth, having a sound competition policy in data-driven industries is an essential component of good governance in the age of big data.

History suggests that competition policy requires an in-depth understanding of business practices and their potential harms and benefits, together with the ability to evaluate efficient market performance -- the production and sale of the highest possible quantity and quality of goods and services at the lowest possible prices. The regulatory objective is straightforward: "the goal of antitrust is to make sure that consumers benefit from the forces of competition" (Shapiro, 2018). Indeed, the objective is to promote competition and market efficiency. In that regard, different degrees of market concentration or data-related barriers to entry could either be a problem that damages consumers and competition, or a natural result of superior business competence within an intensely competitive market environment, and which promotes enhanced product quality and lower prices. In any specific situation, it all comes down to whether the evidence shows the practices in question actually lessen or intensify competition.

While we do not have all the answers to all of the relevant questions, we discuss below the evidence and logic behind several key data-related issues in the digital economy.

[30] How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did, 16 February 2012, https:// www.forbes.com/sites/kashmirhill/2012/02/16/ how-target-figured-out-a-teen-girl-was-pregnantbefore-her-father-did/#6723f52b6668

6.2.1 Do businesses use big data to discriminate against and harm consumers?

In theory, having increasing amounts of information about consumers gives sellers the ability to charge them prices above their costs of production. In the limit, a monopolist might be able to extract all the "consumer surplus," as economists call it, from consumers, charging each of them a price that equals their maximum willingness to pay for the very same, identical product. This is as when an airline charges its flyers wildly different prices for occupying the same seat. Economists call this "perfect price discrimination" though to the non-theoretician there doesn't seem to be anything "perfect" about it.

Consumer surplus in this context refers to the excess that consumers would be willing to pay for a good or service, over and above what they are required to fork over in a given transaction. The "art" of any deal is for consumers to try to obtain the lowest possible prices they have to pay and for sellers to extract from them the highest prices their customers are willing to pay. In a competitive market that price is struck at the point where the price to every buyer is just equal to the marginal or incremental cost to every seller. Competition then greatly constrains the extent to which sellers can raise prices above cost and to charge different prices to differently situated buyers.

As problematic as it sounds, some markets require price discrimination in order for their companies to survive. Public utilities are a common example, in which there is only room for one company to succeed and government regulators allow the monopolist to charge prices well above its marginal cost of production. This prevents the monopolist’s customers from engaging in arbitrage that might otherwise equate prices one to another and to the utilities' marginal costs of production; for example, by stringing up wires between one house or business and another. Without the power to discriminate, it is thought, the monopolist cannot earn enough revenue to survive. Similarly, in the current pandemic, airlines that charge customers different prices for the same seat may need the extra revenues to prevent them from going out of business.

Charging different prices to different consumers in a market can also advance the interests of social welfare, as when theaters, restaurants and numerous other businesses charge lower prices to infants, students, and seniors than to the working population, or when a drug company in an advanced economy sells life-saving drugs at lower prices to citizens living in impoverished regions of the world.

It is important to understand that "price discrimination" does not mean charging different prices for different products. One seat on airline may not be the same as another, even when situated in the same section of an airplane. Each varies in quality according to whether it is purchased for an immediate or long-term use and whether it goes to one ultimate destination or another. Similarly, the exact same seat in a movie theater is worth more at night than in daytime matinees and theaters often price accordingly. Efforts to design, produce, and sell better and more valuable products at different prices that are justified by different benefits to consumers is the stuff of competition, and of what Joseph Schumpeter characterized as "creative destruction":

> "Economists are at long last emerging from the stage in which price competition was all they saw. ***As soon as quality competition and sales effort are admitted into the sacred precincts of theory, the price variable is ousted from its dominant position. ...*** In capitalist reality as distinguished from its textbook picture, it is not that kind of competition which counts but the competition from the new commodity, the new technology, the new source of supply, the new type of organization (the largest-scale unit of control for instance) – competition which commands a decisive cost or quality advantage, and which strikes not at the margins of the profits and the outputs of the existing firms but at their foundations and their very lives. This kind of competition is as much more effective than the other as a bombardment is in comparison with forcing a door, and so much more important that it becomes a matter of comparative indifference whether competition in the ordinary sense functions more or less promptly; ***the powerful lever that in the long run expands output and brings down prices is in any case made of other stuff"*** (Schumpeter, 1962; emphasis added).

Having more information does not automatically lead to price discrimination that is harmful to consumers. In the digital age, the special relationship between business and customers is transforming the nature of competition, just as Schumpeter might have predicted. Today’s producers have unprecedented direct connections to consumers and know much more about what they want from them. Inclusion, i.e., providing goods and services at affordable prices to a larger and larger number of consumers, has become a business priority (Luohan Academy, 2019). In that regard, Ichihashi (2020) shows that digital platforms prefer encouraging information disclosure to the consumers, and therefore have very little incentive to discriminate against them based on price. What’s more: two recent studies "find evidence suggesting that the transparency of the web imposes a constraint on brick-and-mortar retailers' ability to price discriminate across locations ... [suggesting] that as traditional retailers compete more with online retailers, their geographical price dispersion will continue to fall." (Cavallo, 2018, summarizing his research and that of Ater and Ribgi, 2018.) While problematic incidents no doubt exist, there has been little evidence suggesting that using big data to take advantage of consumers has become a pervasive issue.

Price discrimination is not the only concern. "Companies can use big data to exclude low-income and underserved communities from credit and employment opportunities" and this is a concern of consumer protection agencies in the U.S., such as the Federal Trade Commission, which prosecutes unfair trading practices (FTC, 2016). Once again, the increasingly inclusive nature of digital finance in China and

6.2.2 Has big data blocked competition leading to "winner-take-all" market outcomes?

Many of the anti-competitive concerns about big data rest on the perceived strength of the data-feedback loop and “network effects” involving big firms with access to unprecedented amounts of data. While entry barriers naturally vary across industries and over time, some think they are very high in data-driven markets because of sizeable network effects (direct or indirect) and the attendant economies of scale, leaving room for few competitors and a winner-take-all outcome.

Yet in many industries, the reality is far different from the theoretician's hypothesis. Data-driven markets are, in reality, characterized by low entry barriers and fierce competition. Incumbent firms are constantly being forced to protect themselves from potential (and actual) competitors in order to survive. The digital economy offers many examples.

Friendster, originally a "market leader" in the social network industry, was replaced quickly by MySpace, which has now been rendered almost completely obsolete by Facebook. In China, where online consumption now exceeds 25% of all retail sales, competition among Taobao, JD, and Pinduoduo and other players is a typical example of how quickly and mercilessly incumbents can be challenged and displaced by new entrants (Figure 35).[31] While Alibaba is continuing to grow through continuous innovation, its earlier leading position in online sales has not prevented new players from mushrooming all around, with its market share falling by 22 points in the space of the four years from 2015 to 2019. Pinduoduo was able to attract more than four hundred million users while its sales grew by a factor of more than a hundred within three years.

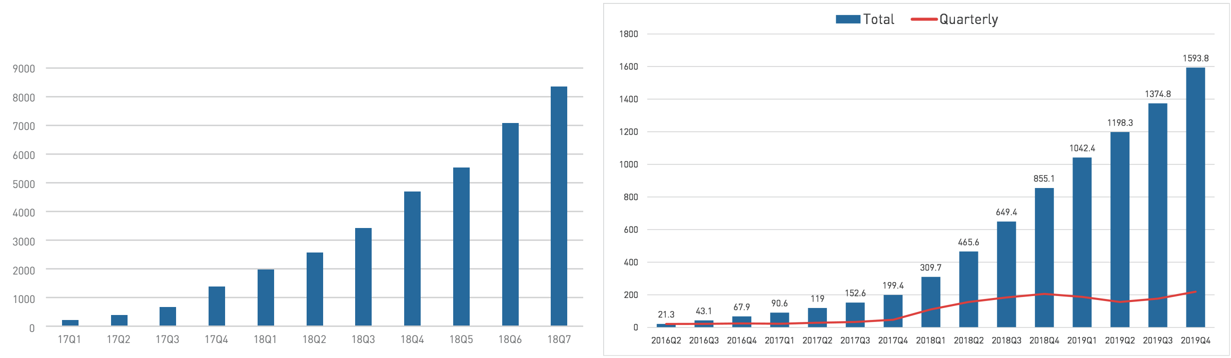
JD.com, with its 17% of China's e-commerce sales and which enjoys the financial backing of America's Google and Walmart giants, recently became "the platform with the largest market share of all channels in the home appliance market."[32] Baidu was China's “dominant” leader in big data and artificial intelligence as of 2010, with a larger market cap than Tencent and Alibaba. It now trails far behind. ByteDance, the mother company of TikTok came out of nowhere to displace Baidu as the market leader in advertising revenue.

The competition between Alipay and WeChat Pay provides yet another example. As the "first-mover" innovator of digital payments in China, Alipay accounted for close to 80% of online payments back in 2014. But by 2019 its market shared had shrunk steadily to 43% as Wechat Pay quickly caught up (Figure 36). Again, in many areas, an early advantage through successful use of big data does not take online platform companies to the land of winner-take-all. Any company that seeks to rest on such laurels is doomed to an early death.

[31] See "Why Can’t Taobao Defeat Pinduoduo," https://seekingalpha.com/instablog/49925729-dongtalk/5264600-why-cant-taobao-defeat-pinduoduo

[32] https://en.wikipedia.org/wiki/List_of_largest_Internet_companies, https://supchina.com/2020/08/07/the-biggest-ecommerce-companies-in-china-a-brief-guide/, and https://equalocean.com/briefing/20200728230002802

Figure 35. New Applications Can Rise Fast



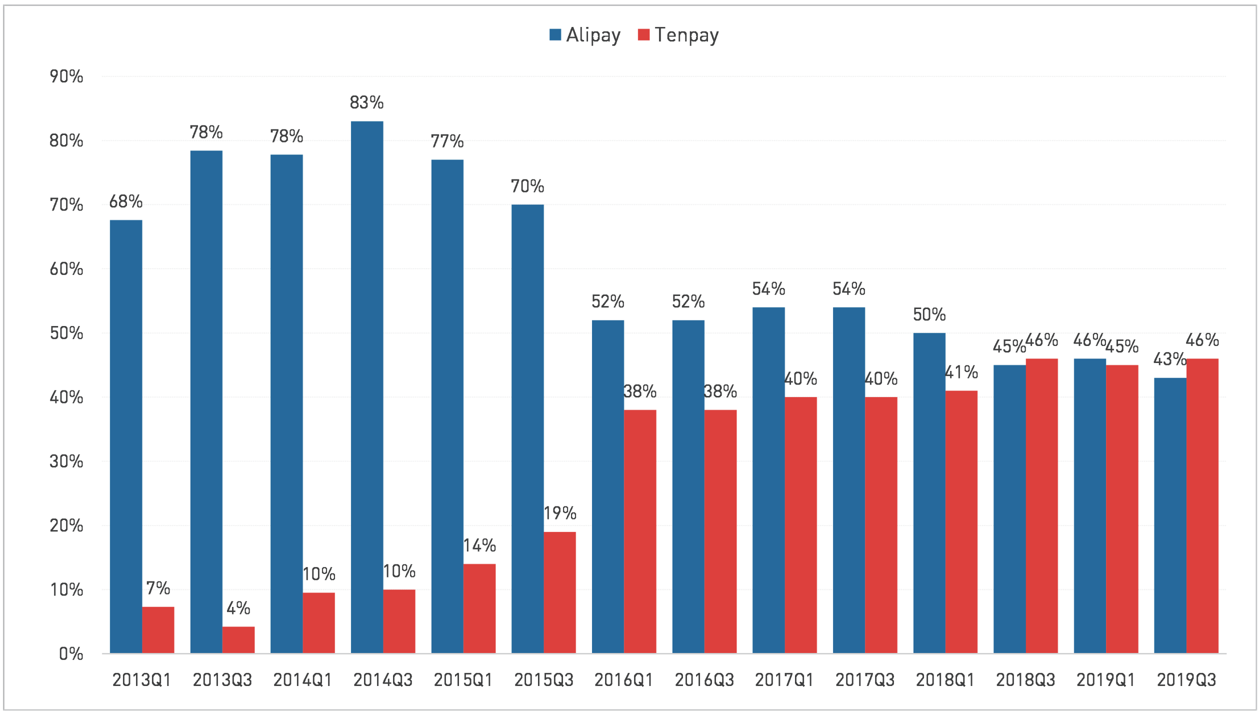
(a) ***Pinduoduo,*** **(b)** ***TikTok First-Time Installs (Global, millions)***

GMV (annual, in hundreds of millions of RMB)

Source: Sensor Tower, organized by Luohan Academy

Note: (1) TikTok was launched in 2016. In 2019 it is the 3rd most downloaded app (behind WhatsApp and Messenger, ahead of Facebook). (2) The installs of TikTok do not include third-party Android downloads in China and other regions.

Figure 36. Market Share in China's Mobile Payment



Source: iResearch; Luohan Academy

In addition to Adam Smith’s observation regarding the pro-competitive effects of expanded trade (see chapter 1), there are several other reasons why big data is far from a position of dominance.

First, the wise and efficient use of data is a necessary but not sufficient condition for success in the digital economy. It is, to be sure, an integral part of the business model. But its increasing importance does not keep Adam Smith’s competition from forcing companies to design, produce, distribute, and relate to customers as they find them in the marketplace for information. The latest data algorithm will not help a digital platform if it can't meet the requirements of the business model.

On the other hand, success in the digital world affords additional opportunities for new entrants to create unforeseen market niches that threaten the success of existing companies. In virtual space, users of digital services can use "multi-homing", allowing users to choose multiple different providers for similar services and to spread their data around the Internet. Once some innovative entrant succeeds in identifying a niche demand that the incumbent does not cover, the entrant can easily squeeze in and quickly catch up with the data-feedback loop and network effects cycle.

Second, the marginal benefits of using data are decreasing with additional volumes of data. For a resource to provide a company with a competitive advantage, it must be inimitable, rare, valuable, and sustainable (Lambrecht and Tucker 2017). Data are often none of these.

Real-time data flows are short-lived. Data have a limited lifespan. Data are being constantly generated and their value declines through time. New entrants are unlikely to be significantly disadvantaged relative to the incumbents because any competitive advantage derived from a specific method of data collection, analysis, or use is ephemeral. Potential competitors do not need to create a data store "equivalent to the size of the incumbent"; instead, they merely need to devise a strategy to accumulate highly relevant and timely data (Schepp and Wambach, 2016). For example, Pinduoduo and TikTok expanded their businesses within just two years, despite the existence of very successful incumbents with huge stores of data.

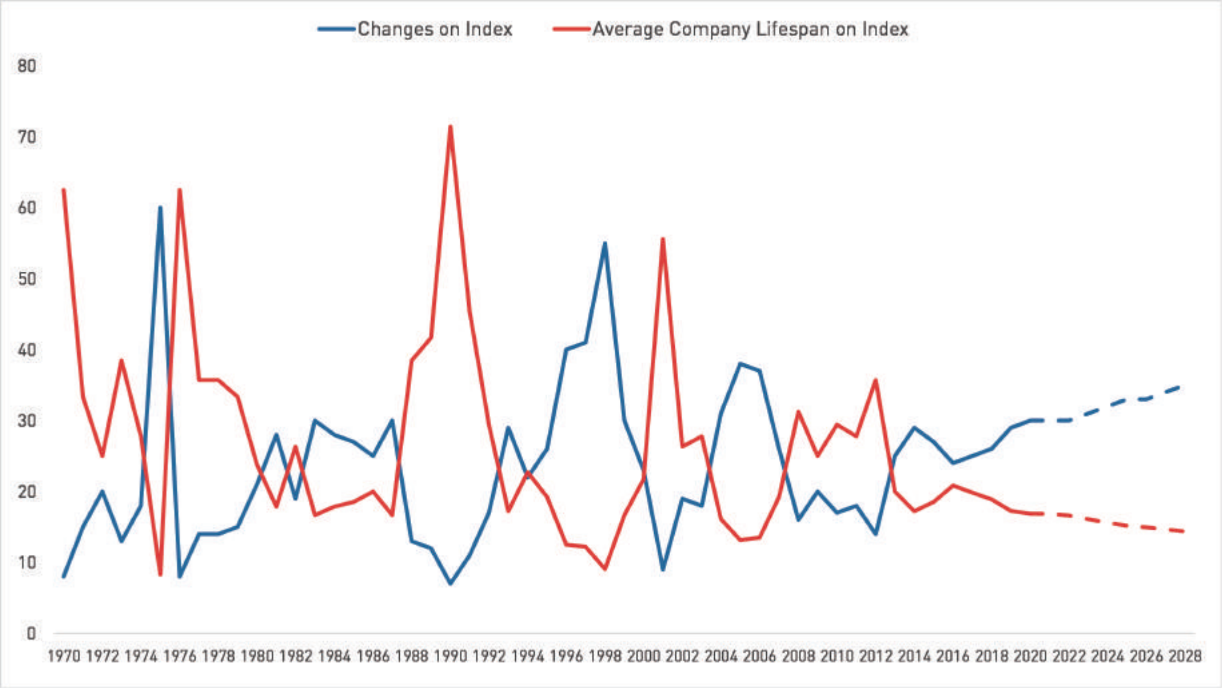
A number of studies show that increasing the volume of data seldom makes a material difference to competitors. For instance, Bajari et al. (2019) use sales data to show that, while having more data about a particular product leads to better forecasts, the marginal value of having additional data falls. Forecasts get better with time, which comes from using the data rather than just having more of it. In addition, increased sales data about other products does not increase forecast accuracy. Chiou and Tucker (2017) find little evidence suggesting that reducing the length of time for which search data are kept (in one example, from 13 to 3 months) appreciably lowers the quality of Internet searches.

As a result and in the words of Shapiro and Varian (1998), market “dominance” in the information industry is fragile and transient: "Hardware and software firms vie for dominance, knowing that today's leading technology or architecture will, more likely than not, be toppled in short order by an upstart with superior technology."

We do not mean to argue that the use of big data can never give rise to monopoly power. Since the collection, compilation, analysis, and use of data are part of the digital business model, their impact will vary from industry to industry and specific situation to specific situation. What is clear is that big data, in and of itself, is not a means to lasting market dominance.

Third, there is ample additional evidence to further support this inference. In the last decade, the average lifespan of companies listed on the S&P 500 index decreased, while the number of entries was increasing, suggesting increasingly vibrant competition (Figure 37). Big data or not, the global business world is entering a stage of increasingly fierce, Schumpeterian competition with shorter business lives even for "dominant" companies that supposedly have locked in impenetrable data advantages.

Figure 37.The Average Lifespan of Companies Listed On the S&P500 Index



Source: S&P Index, projections from QAD Blog.

6.2.3 Do companies use big data to block innovation?

Another important concern of competition policy is whether big data can be used to forestall the entry of innovative competitors. In theory, because big data is a core component of the digital business model, it can be deployed to strengthen a company's current position in order to forestall their entry. Of course, there is concern if the existing competitors are simply better at finding ways to satisfy their customers. The purpose of competition policy is not to prop up inefficient potential (or actual) competitors. Superior efficiency is not a negative, but rather a positive result of competition – especially the kind of competition that has proven to be a gale force for innovation in the digital economy.

China's experience suggests that it is not easy for a "big data company" to "take all" by running other companies out of business, then raising prices to compensate for the initial below-cost pricing. Nor is this generally true in the United States and other world economies. The U.S. Federal Trade Commission's ***Guide on Predatory or Below-Cost Pricing*** says:

> "Can prices ever be 'too low?' The short answer is yes, but not very often. Generally, low prices benefit consumers. Consumers are harmed only if below-cost pricing allows a dominant competitor to knock its rivals out of the market and then raise prices to above-market levels for a substantial time. A firm's independent decision to reduce prices to a level below its own costs does not necessarily injure competition, and, in fact, may simply reflect particularly vigorous competition. Instances of a large firm using low prices to drive smaller competitors out of the market in hopes of raising prices after they leave are rare. This strategy can only be successful if the short-run losses from pricing below cost will be made up for by much higher prices over a longer period of time after competitors leave the market. ***Although the FTC examines claims of predatory pricing carefully, courts, including the Supreme Court, have been skeptical of such claims" (FTC, 2021; emphasis added).***[33]

Claims, especially those of an industry leader's competitors, regarding other suspected anti-competitive practices must also be taken with a few grains of salt. As Ronald Coase (1937) observed, regulators who rely on mathematical economic models often jump to "monopoly" as an explanation for business practices that they do not understand. The FTC's ***Guide on Exclusive Supply or Purchase Agreements*** cautions against concluding that exclusive dealing can only be explained by monopoly:

> "Exclusive contracts can benefit competition in the market by ensuring supply sources or sales outlets, reducing contracting costs, or creating dealer loyalty... [E]xclusive contracts between manufacturers and suppliers, or between manufacturers and dealers, are generally lawful because they improve competition among the brands of different manufacturers (interbrand competition). However, when the firm using exclusive contracts is a monopolist, the focus shifts to whether those contracts impede efforts of new firms to break into the market or of smaller existing firms to expand their presence" (FTC, 2021, emphasis in original).[34]

In an article entitled, "Antitrust and the Winner-Take-All Economy," Alden Abbot, who as FTC General Counsel argues on behalf of the Commission's antitrust and consumer privacy protection litigation before the U.S. Supreme Court,[35] says that America's digital "tech giants bestow huge benefits on American society ... [and that] using antitrust law to attack companies on the basis of size, fairness, or political clout is a recipe for reduced innovation and economic stagnation. . . . Compelling such firms," he says, "to share the source of their advantage is in some tension with the underlying purpose of antitrust law, since it may lessen [their and their rivals'] incentive ... to invest in those economically beneficial facilities. Enforced sharing also requires antitrust courts to act as central planners, identifying the precise price, quantity, and other terms of dealing, a task for which they are ill suited. . . .***Thus, as a general matter the [U.S. Supreme Court] does 'not restrict the long-recognized right of [a] trader or manufacturer engaged in an entirely private business, freely to exercise his own independent discretion as to parties with whom he will deal*** (Abbot, 2018, quoting language from Supreme Court decision, Verizon Wireless v. Trinko, 540 U.S. at 407-08)." (Emphasis added.)

[33] https://www.ftc.gov/tips-advice/competition-guidance/guide-antitrust-laws/single-firm-conduct/predatory-or-below-cost

[34] https://www.ftc.gov/tips-advice/competition-guidance/guide-antitrust-laws/single-firm-conduct/exclusive-supply-or

[35] https://www.ftc.gov/about-ftc/biographies/alden-abbott

Abbot notes that "big, high-profile platforms have two-sided market characteristics [acting] as 'middlemen' that efficiently match two or more sets of individuals (or companies) that want to (or may not want to) deal with each other. ... Two-sided markets display 'network effects,' which means that as the number of participants in the market rises, its value to participants rise, incentivizing more participants to join." He says that these huge network economies are threatened when European and Asian antitrust authorities engage in "intrusive investigations" of the leading firms at the behest of less efficient competitors seeking protection from the leading platforms' competition. ... "[T]he supply of beneficial innovations will slow and consumers will be less well off. What's more, competition will weaken, as the incentive to innovate and compete with market leaders will be reduced. Regulation and public and government favor will substitute for welfare-enhancing goods, services and platform quality." He notes that "the facts bear out this statement. Despite multiple pronouncements by European Union officials that European policy is geared to making Europe a global leader in the digital economy, all of the major digital high-tech platform companies [in the Western world] are American" (Abbot, 2018).

As noted by Furman et al (2019), adoption of big data technologies "can continue to bring benefits to consumers and competition in the form of lower costs for suppliers, better service, better product availability, and an improved customer experience…online platforms can be strong drivers of innovation, and the services they provide to consumers are frequently free at the point of use."

We can think of at least two specific, concrete reasons why the use of data has proven to have been such a powerful driver for competition and innovation – its offspring – in China's economy.

First, Chinese policy makers have not thrown the innovation "baby" out with the privacy protection "bathwater." They have allowed the three "V's" of data to become powerful drivers for production and business model innovation. They have amplified the ability of companies to connect and understand their customers, to make smarter decisions, and to experiment with innovation. In almost every industry where digital technology has played a transformational role -- from education, commerce, medicine and finance, to social media, taxi-hailing, bicycle-sharing, watching videos, and playing video games -- the common denominator has been the development of innovative business models powered by digital technology, with big data replacing traditional business models of industrial organization. The new and diverse models tend to be so innovative that it has become a nearly global phenomenon that the most innovative players are often new entrants into the industry, with little initial capital and other resources and who morph into dominant firms in a very short period of time. ***Innovation, not monopolistic dominance, has become the dominant theme of the digital age.***

There is scarcely any evidence that early movers can block later, innovative entrants. A reasonable perspective for antitrust regulators is probably to start with the view that efficient use of data is necessary for business success in online commerce, and then to analyze the issues case by case, much as is done by economists at the U.S. Federal Trade Commission.

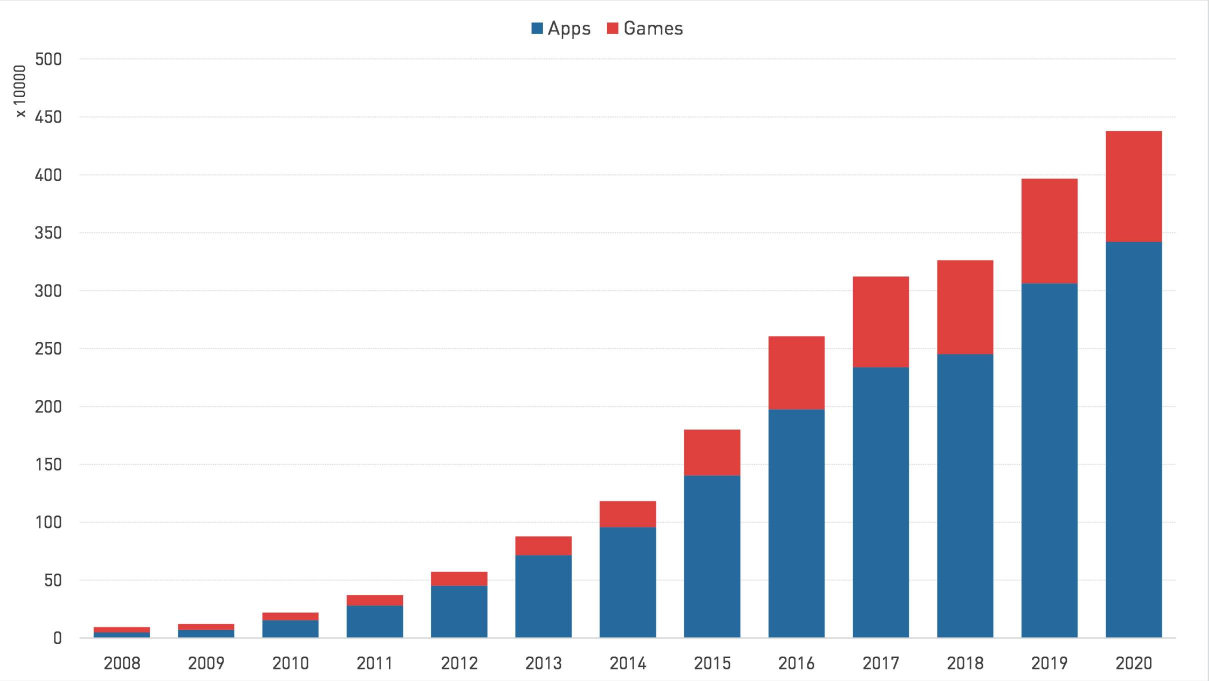
Second, China's and America's policy and business environments have allowed the three "V's" of data to lead to an unprecedented level and scope of collaboration, fostering similar progress in market-wide innovation. Digital platforms have become important promoters of innovation, both within and across business, government, and non-government institutions.

By way of example, since Apple opened its App Store to third-party developers, it has seen wild growth enabling it to strengthen its reputation for providing exceptional user experience. The ecosystem of Apple App Stores creates a competitive marketplace for developers, promoting the innovation of mobile phone software and even new business models built on its collaborators' apps. The symbiotic relationship drives the developers to continuously improve the quality and competitiveness of their services to their, as well as Apple's, benefit. Apple's platform facilities save the developers from routine and often challenging tasks during the distribution process so that they can focus on building up their core business.

Apple designs developer tools, such as TestFlight and App Analytics to help them along. The success of the developers relies heavily on the reviews of users, which financially motivates them to build increasingly better and more competitive apps to meet customers' needs.

Apple announced that the earnings it has paid to app developers who sell digital goods and services through the App Store or within their apps have totaled more than $155 billion worldwide since 2008. According to the Analysis Group (2019), the App Store ecosystem supported $519 billion in billings and sales globally in 2019 alone. As shown in Figure 38, the number of apps available in the App Store has grown from the initial 500 to more than 4 million. These innovative apps have helped to revolutionize the way the world entertains, learns, works, shops, and connects.

Figure 38. Number of Available Apps in the Apple App Store from 2008 to 2020



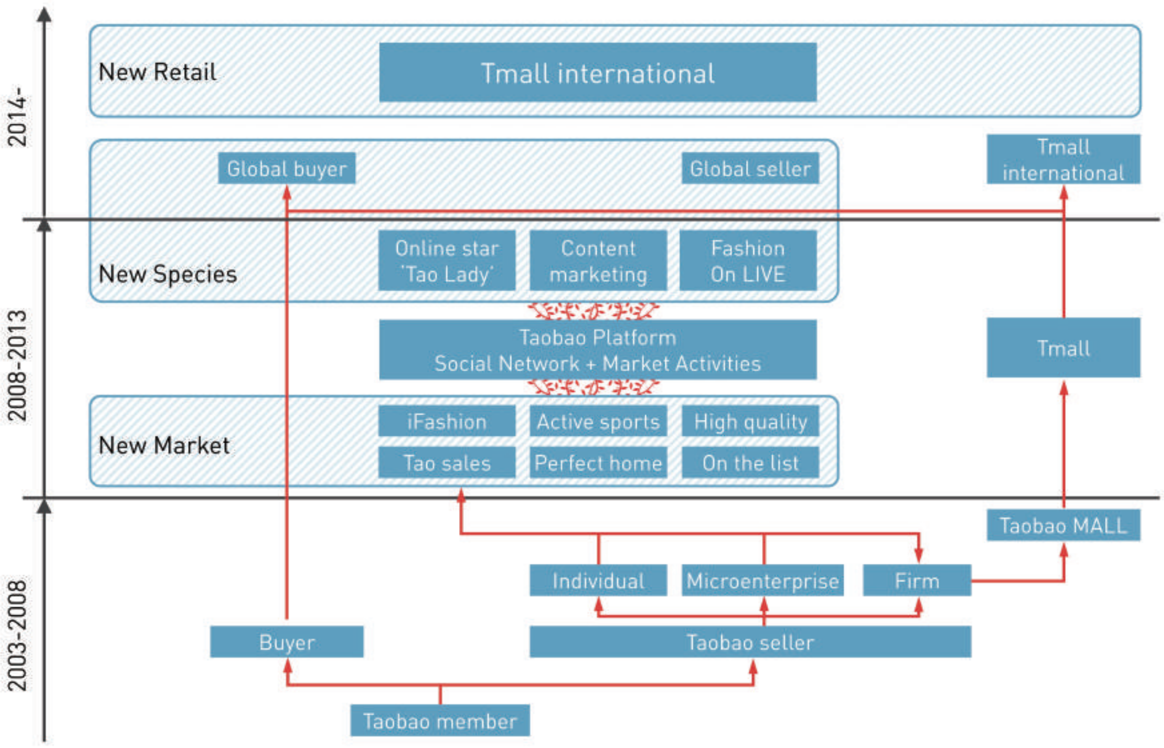
Source: *Statista*

Note: ***The Apple App Store opened on July 10, 2008, via an update to iTunes. This coincided with Apple's launch of the second-generation iPhone 3G which supported mobile apps. Applications in the App Store are only available for iOS devices – the tech industry frequently refers to Apple's ecosystem of devices and operating systems as a "walled garden."***

Another example is supplied from within the Alibaba ecosystem. Alibaba's platforms and those of other e-commerce and fintech companies have created a cornucopia of opportunities for energetic entrepreneurs and their employees. When Alibaba’s Taobao was launched in 2003, individuals were offered the chance to sell their wares through its online website. Sellers were registered as merchants, becoming more like well-established businesses. Niche markets were created in sports equipment, home decorations, global products, and fashion, among others. Many new brands were born. As these markets evolved, they gave birth to many new job categories such as factory sellers, designers, Internet celebrities, specialists, fashion trendsetters, content marketers, customized service providers, and so forth.

Charles Darwin, whose work was inspired by that of the economist, Thomas Robert Malthus (Gordon, 1989), would surely have been fascinated by the rapid evolution of these ecosystems (Figure 39).

Figure 39.The Evolution of Job Descriptions on Taobao



Source: ***Taobao; Luohan Academy***

Similar big data-enabled collaboration is happening in medical innovation as well. McKinsey (2013) reviewed the company profiles and business models of participants in the 2011 and 2012 Health Data Initiative Forum and found that large-scale collaboration based on the big data revolution has created many new species of healthcare innovators.

A typical example is Asthmapolis, which created a GPS-enabled tracker, "Propeller," that monitors inhaler usage by asthmatics. This information is merged with information about known asthma catalysts (for instance, pollen counts in the Northeast and volcanic fog in Hawaii). Knowing more about their patients and the path-breaking findings in the field, physicians can develop personalized treatment plans and spot prevention opportunities. They partner with pharmaceutical companies, MedTech companies, hospitals, healthcare providers, etc., which not only provides instant benefits for all participants but also enables further innovations.

6.2.4 A Vibrant Market is Emerging for the Development of Privacy Protection

Finally, as extensively documented in Chapter 4, there is a growing and intensely competitive market for the protection of privacy, one created by digital platforms themselves. As Bret Swanson, president of Entropy Economics says, "We probably underestimate the natural incentives of firms to protect privacy. Privacy will only become a larger portion of the commercial value proposition" (Swanson, 2019). Or as Maureen Ohlhausen, former Acting Chair of the U.S. Federal Trade Commission, noted in 2016:

> "Concerns about the effects of inaccurate data are certainly legitimate, but policymakers must evaluate such concerns in the larger context of the market and economic forces companies face. Businesses have strong incentives to seek accurate information about consumers, whatever the tool. Indeed, businesses use big data specifically to increase accuracy. ***Our competition expertise tells us that if one company draws incorrect conclusions and misses opportunities, competitors with better analysis will strive to fill the gap"*** (FTC, 2016; emphasis added).

Indeed, digital platforms are furiously competing with one another to find better and better ways to reduce the costs and increase the benefits of providing privacy protection, working towards the day when there may no longer be a daunting tradeoff with the enormous benefits of the digital economy.

Such competition can be suppressed when excessively stringent government regulations divert scarce corporate resources from pursuing these goals. As Swanson observes, "Laws and regulations ... cannot solve every problem, or even most problems. Evolving social norms, more robust institutions, and new privacy-promoting technologies will actually do most of the heavy lifting of protecting our privacy AND promoting data flows" (Swanson, 2019).

Chapter 7.
Concluding Remarks

That information diffusion is a primary driver of economic activity and human prosperity is not a novel insight. What is new is the revolutionary role of digitized information. We are only beginning to understand the enormous potential for big data, making possible an ever-expanding volume, variety, and velocity of communication among buyers and sellers over ever broader and deeper markets. And we are only beginning to grasp its potential for expanding markets across local, regional, and international boundaries, increasing competition and raising productivity and its byproduct: inclusive and sustainable economic growth and prosperity.

As in the past, technological innovation tends to bring challenge as well as opportunity. There are heated, ongoing debates about the value offered by the big data calculus, who should own the data, whether users are getting their fair share of the benefits, to what extent digitized data about individuals is being misused, and how data sharing and digitized businesses should be governed.

In this report, Luohan Academy has sought to offer new insights into several key questions that we believe are critical for a proper understanding and resolution of data-related issues. A key strength of our analysis is that it is not based on some theoretical, mathematical model, but rather on real-world experience within what Ronald Coase called the "molecules of competition." It is based on actual experience with "big data about big data."

To understand what users and subjects value in private information protection, it is simply not enough to rely on what people tell surveys about their attitudes toward privacy. Their preferences are best elicited from their actions and choices where costs as well as benefits are most closely taken into account. Through our study of the "big data calculus," we have learned that users are overwhelmingly willing to exchange personal information for valuable services. They do care about privacy but risks regarding personal information are only one dimension of the issues policy makers should consider. How willingly users share their personal details comes down to whether they trust the service providers, how much information is requested, and the value of the services they receive in return. And the trust they put in service providers depends on the policies those providers put into place in order to protect and preserve the users' privacy and anonymity. Providing certain personal information is often essential for the Internet user to enjoy the benefits of tailored services, especially when buyers are unable to meet sellers face to face in a local offline market. As Hayek made clear, information exchange is both an inevitable and crucial part of all economic activity.

We have also studied why people are willing to exchange information and how its value is affected by its use. The value of data in the digital age can be related to three critical developments: (1) Data sharing has led to unprecedented levels of inclusion and connectivity, transforming the scale and depth of human participation, (2) The availability of big data has led to smarter decision making, especially at SMEs and for individuals who are impoverished or otherwise information-disadvantaged, and (3) information sharing is at the heart of trust-building between sellers and buyers who once were widely separated, increasing market size and intensifying competition within and across greater and greater distances. Our study of big data shows that without a steady flow of personal data, products cannot find suitable customers, and a whole market – the digital market – can be quickly destroyed. Our inquiry has found that economic activity and information exchange within the digital economy are strongly correlated, much more so than ever before. It is striking how digital technology is helping to solve THE economic problem of aggregating and exchanging dispersed information.

While acknowledging the challenges of privacy and data security risks, we have explored how such risks can be effectively and efficiently managed through a middle ground of government and industry self-regulation. With the right design of mechanisms and technologies, it has become increasingly possible to maintain anonymity, collect and share data while avoiding the sharing of personally identifiable information and reducing privacy and security risks, while still allowing data to freely flow. With the right technologies, the benefits of data sharing do not have to conflict with unacceptable risks to privacy. There is a way forward to capture the enormous benefits of big data while mitigating its risks, the goal of efficient and effective privacy protection.

One major issue is data ownership. Giving ownership of data to users who are the subjects of the data may seem like a natural safeguard of privacy. But exclusive ownership would run up against the efficient use of data as a non-rivalry good. In practice, individuals are seldom willing to make the effort of producing and recording data. In the language of economists, the private provision of a public good is generally inefficient. In addition, most people on the street do not have the capacity to mine and create big data for innovation. Data producers -- engineers at information technology firms -- do.

At the Luohan Academy's 2019 conference, a representative from the EU advocated consumer ownership of data, but this paper has shown that individual ownership harms everyone. The Dec. 15, 2020 Luohan Webinar by Zhiguo He similarly showed how data ownership can harm all consumers. Therefore, we would argue that ownership and control of data should at least be shared among data producers and subjects. Perhaps the best way to think about data ownership is as protection for data producers as a whole, including data subjects as "producers," who are thereby given privacy and security rights.

The interaction between data producers, data users, and data subjects is a byproduct of economic and social activity. That is, data are produced, and their value is realized when they are put to use. There is no standalone value of data that is never in use. Once one puts the triangle of data producer, subjects, and use cases together, many misperceptions about the value and ownership of data can be laid to rest.

In this report, we have outlined a simple data analytical framework, the big data calculus, and used it to guide our empirical analysis of how users handle their personal data, where data use generates value, how privacy and security risks can be mitigated, how data are produced stored and accessed, and how the participants in this market might best be governed. We think the evolving regulations of personal data protection are better understood when viewed through the lens of this framework.

Regarding the issue of how data affects competition, we recognize that data can potentially be used as a weapon to reduce competition, but we argue that policy makers should examine case-by-case facts when drawing inferences. In this regard, we have offered evidence on three aspects of this complex issue. First, we have highlighted how business models powered by technology and data have been transformed and innovated by the rise of big data. Second, we have found evidence in China indicating that the use of big data is unleashing vigorous and unprecedented competition. We have found little evidence of price-discrimination or other behaviors that are harming consumers. Third, we find evidence that rather, big data, by rendering possible a more granular matching between suppliers and customers, is promoting innovation as well as rapid entry and subsequent expansion by startups and existing competitors.

We conclude by recommending the following three principles for governing the market for digital data:

Principle 1: Data ownership by data producers (including data subjects as producers) should be predicated on data integrity, anonymity, and especially the protection of personal and societal privacy.

Principle 2: Privacy protection and data security can to a large extent be achieved by combining state-of-the-art technologies and innovative mechanism designs.

Principle 3: Competition and consumer protection analyses of and policy prescriptions for data-driven markets should take into account the documented pro-competitive and pro-consumer benefits of big data along with any potential for anti-competitive and anti-consumer effects in specific markets.

Bibliography

Abbott, A. (2018), Antitrust and the Winner-Take-All Economy, *Legal Memorandum,* No. 224, Heritage Foundation.

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science, 347*(6221), 509–514.

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature,* 54(2), 442–92.

Agrawal, A., Gans, J., & Goldfarb, A. (2018).*Prediction machines: The simple economics of artificial intelligence. Harvard Business Press.*

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics, 84(3),* 488–500.

Anwyl, J. (2011). *Take Polls With a Grain of Salt.* At https://www.edmunds.com/industry-center/analysis/take-polls-with-a-grain-of-salt.html

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science, 355*(6324), 483–485.

Athey, S., Catalini, C., & Tucker, C. (2017). *The Digital Privacy Paradox: Small Money, Small Costs, Small Talk* (No. w23488; p. w23488). National Bureau of Economic Research. https://doi.org/10.3386/w23488

Atkinson, R.D. (2018).*How ICT Can Restore Lagging European Productivity Growth.* Information Technology and innovation Foundation.

Bajari, P., Chernozhukov, V., Hortaçsu, A., & Suzuki, J. (2019).The Impact of Big Data on Firm Performance: An Empirical Investigation. *AEA Papers and Proceedings, 109,* 33–37. https://doi.org/10.1257/pandp.20191000

Banisar, D., & Davies, S. (1999). Global trends in privacy protection: An international survey of privacy, data protection, and surveillance laws and developments. *J. Marshall J. Computer & Info. L., 18, 1.*

Berg, T., Burg, V., Vanjak, A., & Puri, M. (2020). On the rise of FinTechs: Credit Scoring using Digital Footprints. *Review of Financial Studies,* 33(7), 2845–2897.

Blackwell, D. (1953). Equivalent comparisons of experiments. *The annals of mathematical statistics,* 265-272.

Boisot, M., & Canals, A. (2004). Data, information and knowledge: Have we got it right? *Journal of Evolutionary Economics,* 14(1), 43–67.

Cavallo, A. (2018). More Amazon Effects: Online Competition and Pricing Behaviors. Kansas City Federal Reserve Bank.

Carriere-Swallow, Y., & Haksar, V. (2019).The Economics and Implications of Data : An Integrated Perspective. *IMF Working Papers.*

Cate, F. H. (2006). The failure of fair information practice principles. *Consumer Protection in the Age of the Information Economy.*

Chen, L., Huang, Y., Ouyang, S., & Xiong, W. (2020). Data Privacy Paradox and Digitial Demands. Working Paper.

Chen, X., & Michael, K. (2012). Privacy Issues and Solutions in Social Network Sites. *IEEE Technology and Society Magazine, 31*(4), 43–53.https://doi.org/10.1109/MTS.2012.2225674

Chiou, L., & Tucker, C. (2017). *Search Engines and Data Retention: Implications for Privacy and Antitrust* (No. w23815; p. w23815). National Bureau of Economic Research.https://doi.org/10.3386/w23815

Coase, R. (1994). *Essays on Economics and Economists.* The University of Chicago Press.

Coase, R. H. (1937). The Nature of the Firm. *Economica,* 4(16), 386–405.https://doi.org/10.1111/j.1468-0335.1937.tb00002.x

Cœuré, B. (2020). Leveraging technology to support supervision: challenges and collaborative solutions. Speech at the Peterson Institute for International Finance, Financial Statement event series, August 19, 2020.

Culnan, M. J., & Bies, R. J. (2003). Consumer Privacy: Balancing Economic and Justice Consider ations. *Journal of Social Issues, 59(*2), 323–342. https://doi.org/10.1111/1540-4560.00067

Dempsey, J. (2019). Institutionalizing the Concept of Privacy: Global Convergence and Complexity in the Digital Age. Speech at the Conference on Privacy and Data Governance organized by Luohan Academy on March 19-20, 2019, Hangzhou.

Diamond, P. A. (1971). A model of price adjustment. *Journal of Economic Theory, 3*(2), 156–168. https://doi.org/10.1016/0022-0531(71)90013-5

Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research, 17*(1), 61–80.

Equifax Inc., Louis Harris and Associates., & Westin, A. F. (1991). *Equifax-Harris Consumer Privacy Survey.*

Federal Trade Commission. (2016). Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues. *FTC Report.*

Federal Trade Commission. (1998). Privacy Online: A Report to Congress at https://www.ftc.gov /reports/privacy-online-report-congress

Furman, Jason, et al. "Unlocking digital competition: Report of the digital competition expert panel." UK government publication, HM Treasury (2019).

Global Privacy Enforcement Network. (2018). GPEN Sweep 2018: Privacy Accountability.

Goldberg, Samuel and Johnson, Garrett and Shriver, Scott, Regulating Privacy Online: An Economic Evaluation of the GDPR (July 17, 2019). Available at SSRN:https://ssrn.com/abstract=3421731 or http://dx.doi.org/10.2139/ssrn.3421731

Goldfarb, A., & Tucker, C. E. (2011). Privacy regulation and online advertising.Management Science, 57, 57–71.

Goldfarb, A., & Tucker, C. (2012). Shifts in privacy concerns.*American Economic Review, 102*(3), 349–53.

Goldfarb, A., & Tucker, C. (2019). Digital Economics. *Journal of Economic Literature, 57*(1), 3–43. https://doi.org/10.1257/jel.20171452

Gordon, S. Journal of the History of Biology, 22(3), 1989, 437-459.

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review,* 70(3), 393–408.

Gürses, S., Troncoso, C., & Diaz, C. (2011). Engineering privacy by design. *Computers, Privacy & Data Protection, 14*(3), 25.

Hart, O. D. (1988). Incomplete Contracts and the Theory of the Firm. *Journal of Law, Economics, & Organization, 4(*1), 119–139.

Hart, O., & Moore, J. (1988). Incomplete contracts and renegotiation. *Econometrica: Journal of the Econometric Society,* 755–785.

Hartzog, W. (2016). The Inadequate, Invaluable Fair Information Practices. *Md. L. Rev., 76,* 952.

Hau, H., Huang, Y., Shan, H., & Sheng, Z. (2018). FinTech Credit, Financial Inclusion and *Entrepreneurial Growth. Working Paper.*

Hayek, F. A. (1945). The Use of Knowledge in Society. *The American Economic Review, 35*(4), 519–530. JSTOR.

Hirshleifer, J. (1980), Privacy, Its Origin and Future.*Journal of Legal Studies, 9*(4), 649-64.

Hoepman, J.-H. (2014). Privacy design strategies.*I FIP International Information Security Conference,* 446–459.

Holdren, & Lander. (2014). *Big Data and Privacy: A Technological Perspective.* PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY.

Holmström, B. (1979). Moral Hazard and Observability. *The Bell Journal of Economics, 10*(1), 74–91. JSTOR. https://doi.org/10.2307/3003320

Holmström, B. (1982). Moral hazard in teams. *The Bell Journal of Economics,* 324–340.

Holmström, B. (2018). *Keynote Speech at Toulouse School of Economics.*

Hoofnagle, C. J., Soltani, A., Good, N., Wambach, D. J., & Ayenson, M. D. (2012). Behavioral Advertising: The Offer You Cannot Refuse. *Harvard Law & Policy Review, 6,* 273.

Ichihashi, S. (2020). Online Privacy and Information Disclosure by Consumers. *American Economic Review, 110*(2), 569–595. https://doi.org/10.1257/aer.20181052

Johnson, G. A., Shriver, S. K., & Du, S. (2020). Consumer privacy choice in online advertising: Who opts out and at what cost to industry? Marketing Science, 39(1), 33–51.

Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the Economics of Data. American Economic Review, 110(9), 2819-58.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. American Psychologist, 39(4), 341–350. https://doi.org/10.1037/0003-066X.39.4.341

Kitchin, R. (2014).*The data revolution: Big data, open data, data infrastructures and their consequences.* Sage.

Kummer, M., & Schulte, P. (2019). When Private Information Settles the Bill: Money and Privacy in Google's Market for Smartphone Applications. *Management Science, 65*(8), 3470–3494. https://doi.org/10.1287/mnsc.2018.3132

Layton, R.(2019) *Should Online Privacy Protection Be Based on Trust or Control?* American Enterprise Institute.

Layton, R. (2019A) *Seven Virtues of Data Privacy and Protection.* American Enterprise Institute.

Layton,R. (2019B) *Seven Deadly Sins of the Privacy and Data Protection Debate.* American Enterprise Institute.

Lambrecht, A., & Tucker, C. E. (2017). Can Big Data Protect a Firm from Competition? *Antitrust Chronicle, 1*(12), 17.

Landau, S. (2015). Control use of data to protect privacy. *Science, 347*(6221), 504–506.

Laufer, R. S., & Wolfe, M. (1977). Privacy as a Concept and a Social Issue: A Multidimensional Developmental Theory. *Journal of Social Issues, 33*(3), 22–42. https://doi.org/10.1111/j.1540-4560.1977.tb01880.x

Luohan Academy (2019). *Digital Technology and Inclusive Growth.* Luohan Academy.

McGann, J.G. (2018). 2018 *Global Go to Think Tanks Report and Policy Advice.* University of Pennsylvania, 15.

Martin, K.D. & Murphy, P.E. (2016) *The Role of Data Privacy in Marketing.* Journal of theAcademy of Marketing, p. ?

Maskin, E. (2008). Mechanism Design: How to Implement Social Goals. In *Les Prix Nobel 2007. Nobel Foundation.*

McDonald, A., & Cranor, L. F. (2010).*Beliefs and Behaviors: Internet Users' Understanding of Behavioral Advertising.*

Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research, 6*(1), 58–73.

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life.* Stanford University Press.

North, D. C. (1990). *Institutions, institutional change, and economic performance.* Cambridge University Press.

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. *The New England Journal of Medicine, 375*(13), 1216.

OECD.(2015).IndustrySelfRegulation:RoleandUseinSupportingConsumerInterests. https://doi.org/10.1787/20716826

Organisation for Economic Co-operation and Development (Ed.). (1981). *Guidelines on the protec tion of privacy and transborder flows of personal data.* Organisation for Economic Co-op eration and Development; OECD Publications and Information Center.

Pavlou, P. A. (2011) State of the Information Privacy Literature:Where are We Now and Where Should We Go? *MIS Quarterly. 35*(4), 977-988.

Phelps, E. S. (Ed.). (1970).*Microeconomic foundations of employment and inflation theory.* Norton.

Pissarides, C. A. (2000). *Equilibrium unemployment theory* (2nd ed). MIT Press.

Pissarides, Christopher A. (2009). The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer? *Econometrica, 77*(5), 1339–1369. https://doi.org/10.3982/ECTA7562

Reinsel, D., Gantz, J., & Rydning, J. (2018).*Data age 2025: The digitization of the world from edge to core.*

Romer, P. (1990). Endogenous Technological Change. *Journal of Political Economy, 98*(5), S71-102.

Romer, P. (2018). *On the Possibility of Progress* (Issues 2018–4). Nobel Prize Committee. https://EconPapers.repec.org/RePEc:ris:nobelp:2018_004

Roth, A. E. (2018). Marketplaces, Markets, and Market Design. *American Economic Review, 108*(7), 1609–1658. https://doi.org/10.1257/aer.108.7.1609

Rubinstein, I. S., & Good, N. (2013). Privacy by design: A counterfactual analysis of Google and Facebook privacy incidents. *Berkeley Tech. LJ, 28,* 1333.

Schepp, N.-P., & Wambach, A. (2016). On big data and its relevance for market power assessment. *Journal of European Competition Law & Practice, 7*(2), 120–124.

Schumpeter, *Capitalism, Socialism, and Democracy* (1962).

Shapiro, C. (2018). Antitrust in a Time of Populism. *International Journal of Industrial Organiza tion, 61,* 714–748.

Shapiro, C., & Varian, H. R. (1999). *Information rules: A strategic guide to the network economy.* Harvard Business School Press.

Singh, S. (1999). *The code book* (Vol. 7). Doubleday New York.

Smith, H. J., Dinev, T., & Xu, H. (2011).Information privacy research: An interdisciplinary review. *MIS Quarterly,* 989–1015.

Smith, Mi. D., Bailey, J., & Brynjolfsson, E. (1999). Understanding Digital Markets: Review and Assessment. In E. Brynjolfsson & B. Kahin (Eds.), *Understanding the Digital Economy.* MIT Press.

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics,* 355–374.

Spence, M. (1974). Competitive and optimal responses to signals: An analysis of efficiency and distribution. *Journal of Economic Theory, 7*(3), 296–332.

Stigler, G.J. (1984), An Introduction to Privacy in Economics and Politics, *Journal of Legal Studies, 9*(4), 623-644.

Stigler, G. J. (1961). The Economics of Information. *Journal of Political Economy, 69*(3), 213–225. https://doi.org/10.1086/258464

Stigler, G. J. (1962). Information in the Labor Market. *Journal of Political Economy, 70*(5, Part 2), 94–105. https://doi.org/10.1086/258727

Stiglitz, J. (1974). Incentives and Risk Sharing in Sharecropping. *Review of Economic Studies, 41*(2), 219–255.

Sun, T., Yuan, Z., Li, C., Zhang, K., & Xu, J. (2020). The Value of Personal Data in Internet Com merce: A High-Stake Field Experiment on Data Regulation Policy. *Working Paper.* Avail able at SSRN: https://ssrn.com/abstract=3566758

Tadelis, S. (2002). The Market for Reputations as an Incentive Mechanism. *Journal of Political Economy, 110*(4), 854–882. JSTOR. https://doi.org/10.1086/340781

Tews, S. *Privacy and Europe's Data Protection Law: Problems and Implications for the U.S.* American Enterprise Institute.

UNCTAD, S. (2019). *Unctad Stat Data Center.* World statistical database.

USITC. (2014). *Digital Trade in the U.S.* and Global Economies. U.S.

Veldkamp, L., & Chung, C. (2019, October). Data and the aggregate economy. In Annual Meeting Plenary (No. 2019-1). Society for Economic Dynamics.

Verizon. (2015). *Data Breach Investigations Report.*

Volio, F. (1981). Legal personality, privacy and the family. *The International Bill of Rights: The Cov enant on Civil and Political Rights, 185.*

Warren, S.D. and Brandeis, L.D. (1890), The Right to Privacy, *Harvard Law Review, 4*(5), 193-220.

Westin, A. F. (1968). Privacy and freedom. *Washington and Lee Law Review, 25*(1), 166.

World Bank Group. (2016). World development report 2016: Digital dividends. *The World Bank.*

Xu, H., Teo, H.-H., Tan, B. C., & Agarwal, R. (2009). The role of push-pull technology in privacy calculus: The case of location-based services. *Journal of Management Information Systems, 26*(3), 135–174.

Yao, A. C. (1982). Protocols for secure computations. *23rd Annual Symposium on Foundations of Computer Science (Sfcs 1982),* 160–164.

© 2021 The Luohan Academy
Z Space, No. 556 Xixi Road, Hangzhou, China
www.luohanacademy.com

This work is a product of Luohan Academy (the "Academy"). Unless otherwise specifically stated, the findings, interpretations, and conclusions expressed in this work are solely those of the authors listed, and do not necessarily reflect the views of any of the Academy's affiliated entities or their officers or directors or any advisers and researchers who may have otherwise contributed to this work.

The Academy does not guarantee the accuracy of any data cited, used or otherwise included in this work. Any information shown on any map or graph in this work do not imply any judgment on the part of the Academy concerning the legal status of any territory.

Nothing herein shall constitute or be considered to be a limitation upon or waiver of the rights, privileges and immunities of the Academy, all of which are specifically reserved.

Data Protection Disclaimer

Luohan Academy does not use any personal data in its research or other related activities. It has instituted policies and procedures that require the data it collects or receives for research purposes is fully anonymized, meaning the Academy does not, at any time, access any data that would identify, or allow it to identify, any specific individuals. Any data points attributed to individuals included in any publication based on the Academy's data will only reflect fully anonymized, aggregate and/or scaled information based on random samples.

The Academy also places restrictions on the use of any data that it collects or receives for research purposes. It has put in place data usage protocols for its researchers and other members, requiring those provided with data access to undergo rigorous background checks and execute confidentiality agreements, and providing training on the proper use of data and data privacy rules. Researchers and others who access data provided by the Academy and/or its affiliates are contractually or otherwise obligated to use the data solely for approved research/internal discussions. They are also contractually obligated not to attempt to re-identify any individual represented in the data.

The Academy stores its data on a secure server which can be accessed only when following strict security procedures. The data cannot be exported from the system being used by the Academy to any external servers or email addresses. Data storage measures implemented and maintained by the Academy comply with strict standards for the monitoring and security of data.

LUOHAN ACADEMY

UNDERSTANDING BIG DATA

DATA CALCULUS IN THE DIGITAL ERA