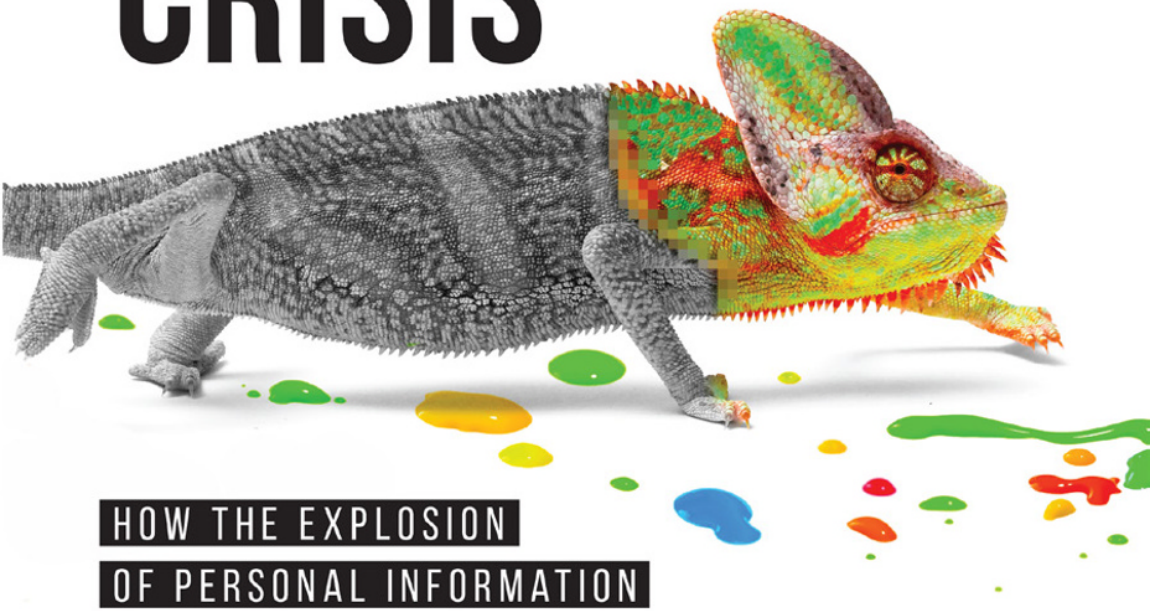
ROOLY ELIEZEROV

THE DIGITAL IDENTITY CRISIS



HOW THE EXPLOSION OF PERSONAL INFORMATION IS TRANSFORMING TECHNOLOGY, BUSINESS, AND SOCIETY

WILEY

The Digital Identity Crisis:

How The Explosion of Personal Information Is Transforming Technology, Business, and Society

by Rooly Eliezerov

WILEY

Publisher's Acknowledgments

For general information on our other products and services, or how to create a custom book for your business or organization, please contact our Business Development Department in the U.S. at 877-409-4177, contact info@dummies.biz, or visit www.wiley.com/go/custompub. Some of the people who helped bring this book to market include the following:

Project Manager and Development Editor: Chad R. Sievers

Executive Editor: Katie Mohr

Editorial Manager: Rev Mengle

Business Development Representative: Karen Hattan

Custom Publishing Project Specialist: Michael Sullivan

Production Editor: Tamilmani Varadharaj

The Digital Identity Crisis

Published by
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

ISBN 978-1-119-47985-7 (pbk)

ISBN 978-1-119-47987-1 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Author's Acknowledgments

I want to thank many friends and colleagues who took part in bringing this book to light: first and foremost to Rebecca Rachmany for her research, Cheryl Dumesnil for content editing, and Chad Sievers for copyediting.

A significant part of the book resulted from discussions and interviews with the following people, who contributed valuable information and ideas: Dan Ariely, Nat Sakimura, Michael Eisenberg, Martin Kuppinger, Hanan Levin, Maria Macocinschi, Assaf Mischari, Daniel Landau, Eyal Magen, Eran Kutner, Moran Cerf, Rashmi Vittal, Kevin Hobbs, Ahnjili ZhuParris, Jochen Scherschmidt, Menny Barzilay, Samit Singh, Dondrey Taylor, Fatemeh Khatibloo, Rotem Hermon, Gonzalo Higueras, Taylor Vallone, Craig Ferrara, Brennan Wright, Marcel Neumann, and Kevin Spier. A special thank you to Colin Wallis.

I also want to acknowledge those who invested their time in reviewing the work and offering thoughtful feedback: Patrick Salyer, Oded Nir, Guy Kronental, Nadav Gur, Mike Langberg, Dave Scott, Yaara Tsory, Adili Nir, Yves Chtepenko, Adrian Nash, Andrew Bud, Ashish Jain, Yaron Gur Ari, Oren Evron, Atiya Davidson, and Joe Dinucci.

To Mom

Contents

Foreword ix

Introduction 1

Where I Come from 1

How This Book Is Organized 2

1 Ownership of Identity 5

Map of Interests 6

Government 6

Businesses 7

What Could Possibly Go Wrong? 8

Overcollecting and Oversharing 10

Data Sharing: It's Complicated 11

What Does Ownership Mean? 12

Your identifier 12

Your data 13

Self-Sovereign Identity 14

Accuracy, quantity, and expiration 16

Where could SSI go? 18

Reputation and Identity 18

How Many of You Are There? 20

Delegation 22

When Will It Happen? 23

2 Privacy 26

What Do You Have to Hide? 26

What Is Data Privacy? 27

Recording Your Private Lives 29
The Mega Players 32
You Aren't as Anonymous as You Think. 33
Control 35
The evolution of consent. 35
How consent works 36
The business end of consent. 39
Who cares? 44
Sharing Data with the People You Know 45

3 Protecting Your Identity 50
Who Are You? 51
It used to be simple. 52
A digital identity 53
A dead man walks into a bank 54
Government, security, and identity 56
Crime and Identity 56
What's it worth? 58
Phishing for data 59
Who protects you? 60
A hacker's ROI 60
High volume, low-cost hacking. 61
Scaled-up exposé 62
Authentication 63
The death of passwords 63
Beyond passwords 66
Could biometrics replace passwords? 68
The Future of Authentication Is Like the Past of Authentication . . .70

4 The Value of Relationships 76

The Wild Frontier, Then and Now 77

Get Creative: Using Customer Data 79

Value → Trust → Exchange 81

Progressive identity 83

Contextual personalization 85

Challenges in Customer Identity 86

How Smart Companies Handle Identities and Profiles 87

5 Identity Data 101 91

The Evolution of Data Use 92

How Far Should Personalization Go?. 93

Not All Data Is Born Equal 94

Identity Types 95

Unverified 96

Verified 96

Federated 96

Reputation-based identities 98

The shared accounts challenge remains 98

Keeping It Appropriate 99

The Creepy Factor 100

What Do Customers Want? 104

6 The Search for a Better Self 107

Learning About Yourself 107

How Do You Decide? 110

The Three Considerations of Personal Data 111

What do you want to know? 111

What do you want to share?. 112

What impact can this data make? 113

Introducing the Quantified Self 113
Knowing yourself, inside and out 114
The technology 114
The future 115
What Can Be Measured? What Can Be Improved? 116
Can data make you happy? 117
Measuring lifestyle 118
Challenges 119
Virtual-World Data, Real-World You 121
Remembering Who You Are 122
Making Better Decisions 124

7 Identity and Artificial Intelligence 126
Who Is Making Your Decisions? 126
What Isn't Known 128
Artificial Intelligence Enters the Picture 129
How Far Will AI Go? 132
Your Agents. 133
Challenges with AI 135
Computers don't have common sense 135
Computers can't make a judgment call 135
AI decision-making is too complex to understand 136
AI presents a unique security risk 139
How Much Will AI Change Society? 139
Who Is Behind the Computers? 141
You Will Delegate 141

Epilogue 144
How Will the Digitalization of Identity Change Us? 144
But There Is a Problem Here 146

Foreword

I am truly a creature of the digital age. I sent my first Internet email from Stanford to a friend at another university in 1978, and several years later I joined the emerging Silicon Valley workforce at a time when personal computers were just appearing and changing everything about how everyday people knew and interacted with technology. Later, I helped to pioneer some of the very first "big data" systems and analytic tools for business people growing hungry for more and better insights about their customers. And in a contribution about which I am somewhat less proud, I helped to build one of the first email marketing systems, starting a trickle of unsolicited commercial email that led to today's unrelenting tsunami of spam.

But I must admit, through these formative years of the connected world we all now share, I didn't give very much thought to my online identity. I didn't really understand that my name and contact information, my online behavior, and my preferences and interests were all tiny precious gems being gathered into an enormous body of digital information that would come to be bought, sold, and ubiquitously exploited by philanthropists, brands, politicians, con men, and thieves. I didn't think much about who had my personal data, how they got hold of it, where they stored it, and what they did with it.

In the intervening years, the scales have fallen from my eyes, and I now really see the essential value of my digital identity. I now understand that respectable businesspeople, organized criminals, and even rogue nation states are in a full blown frenzy to buy or steal my information to gain the slightest edge on their competition or enemies. And of course, billions of my fellow digital citizens around the world have begun to arrive at this same understanding.

We now appreciate that our digital identity is, in a very real sense, essential to how we live our lives. Our everyday activities generate more and more data about who we are and what we do, and we continuously distribute it to all sorts of organizations through the devices we use—our phones, coffee machines, cars, and even the places we live. As a result, in many ways our world is getting smarter and ever more tuned to our desires and preferences. Looking forward,

harnessing our digital information for good purpose will allow us to be more productive, to be better informed, and to create more valuable business relationships and more intimate personal relationships with each other.

But at the same time, this plethora of data we generate about ourselves raises questions that grow in significance every day. Who owns this data? How private is our data? Is it protected well? We collectively spend billions every year in an attempt to secure those precious gems and factoids about our online selves and to battle spam and new kinds of unsolicited digital contact. As we have come to know and understand this battle for our digital identity, we have also come to fear the results. A recent Gallup poll found that more Americans fear identity theft than fear terrorism or a global pandemic like Ebola. And we have begun to demand that our governments do something to protect our digital identities, resulting in regulations like CAN-SPAM and the expansive new GDPR regime put forward by the European Union.

We are at a critical tipping point. Do we succeed at protecting our digital identities, building and maintaining trust between consumers and the businesses and organizations that collect and process this data? Can we create the exciting future that a more personal, better adapted, and smarter environment can produce for all of us? Or does complexity and fear cause us to retreat from the amazing opportunity that may lie ahead if our digital identities are managed and used well? It's in this context that I welcome *The Digital Identity Crisis* by Rooly Eliezerov. It's an important resource as we all seek to be better informed and more able to meet one of the most important issues of our time.

In three sections, Rooly first introduces the topic of our digital identities, why we all should care about them, and yes, why we should also fear that they are under attack. Next, he looks at the challenge from the point of view of the entities that gather, store, and manage our personal information, and offers practical solutions for not only how organizations can achieve better security and privacy, but also how these same best practices can lead to more personalized products and services and greater customer intimacy. Finally, he looks forward, because the forty-year journey I alluded to earlier is just the

lead-up to a near-term future populated by speaking digital assistants and near-spooky artificial intelligence—all of which will be powered, guided, and shaped by the personal information stored about each and every one of us. Rooly ends the book with an intriguing epilogue that reflects a philosophical perspective about the main ethical conflict we will be facing as individuals.

I first had the pleasure to meet Rooly some years ago when I joined the Board of Directors of Gigya, Inc., the company Rooly co-founded, where he was leading the development of a new kind of software platform dedicated specifically to securing customer identity information and empowering customers to take control of their online identities.

At the time I joined Gigya, I was CEO of Marketo, a marketing technology company where we securely stored and managed customer data about hundreds of millions of individuals. There, I was often asked, "What keeps you up at night?" and I really had only one answer: a data breach and a violation of the trust our customers had put in us. So I have lived this topic up close and very personally, understand its complexity, and realize firsthand how important it is to have new expert voices helping us all see the way forward to protect our online digital identities.

I can't think of a better expert voice than Rooly in *The Digital Identity Crisis.* Rooly is a deep thinker and experienced entrepreneur who has worked on this topic for more than a decade, and who has broken real new ground in building practical software solutions to help companies building trusted relationships with their customers. At the same time, Rooly brings a human voice to the topic, creating a book that both experienced business professionals and the everyday owners of our digital identities will find compelling.

Phil Fernandez
Palo Alto, California

Introduction

The moment people were able to create online accounts, their individual identities began to extend exponentially into the digital world. Now, with every site you visit, app you use, and soon every machine you operate, you're creating and sharing aspects of your identity. In one way or another, it seems you're perpetually answering the question, "Who are you?"

Whether you're commenting on a social network, buying an item online, accessing a computer system at work, or using a nutrition app, you're recording bits of who you are. This digitization of identity is transforming the ways human beings conduct business, govern, and manage their daily lives.

All of this is also raising a rather disturbing question: "Where are you?" The fracturing of your online identity has reached crisis proportions. You don't know where this data is stored, who has access to it, or how they're using it. The businesses you interact with, too, are scrambling for solutions to manage customer data respectfully and in accordance with newly evolving regulations.

Inarguably, we're starting to see the value digital tools offer—from personalized content and products, to tailor-made medical care, to enhanced decision-making capabilities. But as the digital revolution marches forward, it's important to examine the opportunities and implications of a digitized identity. How will these advancements affect you—for better and possibly for worse—in the long run?

This is the territory I'll explore in *The Digital Identity Crisis.*

Where I Come from

My career path began in architecture, a discipline that I believed would engage both my analytical mind and my creative expression. But a year into my practice, when my drawings were going to become real buildings, I began to feel limited by architectural design. At best,

my designs were based on informed guesses about how people would use my buildings. Once the structures were complete, I would never be able to adjust the designs to accommodate the complex realities of the people inhabiting them.

I still wanted to create spaces for people, but in a dynamic design environment, one that more fully addresses human complexity. So what better place to land than an Internet startup? In the digital world, I could engage in agile design, taking into account the massive amounts of information people generate, testing and observing the results of my efforts, and improving my designs with each iteration. It didn't take me long to recognize that I was in my element, which led me to found Gigya with Eyal Magen and Eran Kutner.

Gigya started in 2006, when social networks began expanding identity into the digital space. Our company has seen several iterations, responding to the evolving needs of businesses engaging with digital identities. Now, Gigya manages more than a billion customer identities for hundreds of enterprises around the world. As I was finishing this book, we announced the acquisition of Gigya by SAP, one of the world's leading enterprise software companies.

From my vantage point at Gigya, over the past decade I have seen broad-reaching changes in the ways people share their personal data, the way businesses engage their customers online, how governments regulate identity data, and the extent to which data-based technologies transform people's lives.

This book explores the evolution, challenges, impact, and future of digital identity, from the perspectives of individual users and the entities with which they interact.

How This Book Is Organized

The book is organized in three parts.

Part I addresses the three main issues that people struggle with as their identities move to the virtual space: ownership, privacy, and security.

Chapter 1 begins with the idea of digital identity ownership. You spend countless hours online, generating enormous amounts of data about yourself as you extract value from the websites and apps you engage. Who owns this data after you've shared it? Ideally, you want to own and control your personal data, but first you need to understand what data ownership and control entail. You've arrived at a time when true identity ownership, or self-sovereign identity (SSI), is feasible, but what will it take to make SSI a reality?

Chapter 2 examines your rights to know where your data resides and to control its use by the entities that store it. What are the privacy implications of having your data stored in so many different places? What platforms are now evolving to help organizations manage privacy in new ways? What are the implications for the businesses and other entities that store your data in an era of government regulation?

Chapter 3 investigates the ways in which your identity is under attack today. Technology is changing the ways you prove who you are and what authorizations and permissions you have. As digital identities become more valuable, they are more attractive to cybercriminals. What kinds of threats do you face? How do companies secure your identity? Passwords aren't the best way to identify yourself, so what other forms of authentication are in use and in development today? Will they make you safer? Safe enough?

Part II explores digital identity from the point of view of the entities—mostly businesses, but also governments and other institutions—that provide online tools and services. These organizations are just beginning to understand the opportunity that digital culture is opening: the potential for creating long-lasting online relationships with their customers.

Chapter 4 looks into the value of the data collected by commercial entities and how it enables companies to create deeper relationships with consumers. Unlike a billboard advertisement or even a coupon code indicating a consumer's geography, progressive data collection—when plotted carefully—offers a tremendous opportunity to develop a two-way relationship that evolves from initial engagement to brand loyalty to expansion across other owned brands.

Chapter 5 examines issues businesses face while handling customer identities. Here I explore the evolution of consumer data usage, the costs and values of collecting certain kinds of consumer data, and the challenges and solutions for appropriate usage of customer data.

Part III peers further into the future. Where is digital identity headed, and how might it change your life?

Chapter 6 discusses the developing relationship you have with your own data. With thousands of people worldwide tracking everything from their gut bacteria to the pollution levels as they walk through town, they now generate and consume unprecedented amounts of data. Precise records of communications, online behaviors, and many other data streams are now available. Paired with new technologies, your identity data can help you make healthier decisions and lead happier lives. But is it all upside? What are you giving up when you let data analytics drive your lives?

Chapter 7 looks at artificial intelligence. As the systems you use become more intelligent and better at learning your habits, you can begin to delegate your authority to automated solutions. Websites already ask your political views and then suggest what candidate or party best represents you. What's next? Currently businesses influence your decisions, using data to understand your tastes and preferences and to show you more of what they think you might like. To what degree are AI and machine-learning solutions influencing your choices? Do you want these algorithms to make recommendations or even to make decisions for you? What do you gain when you delegate control to machines? What are the costs?

1

Ownership of Identity

You don't need to be an expert in data security, privacy, or identity management to recognize that today's culture is embroiled in a digital identity crisis. The symptoms abound, from high-profile data breaches to fractured relationships between consumers and the businesses that serve them. Most fundamentally, though, this crisis is played out in your individual relationships with your own digital identity.

In the digital world, governments, banks, healthcare organizations, and commercial entities, from phone companies to Internet players such as Google and Facebook, perpetually collect, use, and share parts of your identity, for your benefit and for theirs. This ubiquitous collection and sharing of people's identity data, which often happens unbeknownst to them, has resulted in a fragmentation of their identities. You don't know where much of this data resides, nor do you know who has the right to access it, store it, use it, or delete it.

As a result, in some specific ways, you don't control your identity. This lack of control—or ownership—is a major contributor to the digital identity crisis, because it creates an environment of uncertainty, mistrust, and fear. But technologies under development today could enable people to take back ownership of their identities.

Imagine a scenario in which all people become global citizens, not registered to a central government at birth, but connected to an ever-enriching digital identity. What if their reputations, not government-issued documents, authorized them to enter, live in, and work in a specific country? What if their data weren't scattered across

countless systems, but stored in a decentralized system that gave them control over their identity?

Before I talk about such future possibilities, allow me to take a closer look at the state of today's digital identities.

Map of Interests

Yes, a crisis is at hand. But rather than getting carried away by blockbuster movie-worthy conspiracy theories about corporations with unrecognized three letter names, surveillance, and data manipulation, step back and ask two important questions, "Who is collecting your data?" and "What do they want with it?"

The two main players in the area of data ownership are government and businesses. The following sections take a closer look at each.

Government

In most countries, ownership of your identity is turned over to a central authority the moment your life begins, when your parents register your birth with the government. Parents might not think about the fact that they have—on your behalf—accepted the country of your birth as the authority over your identity, but from that moment on, throughout your life you'll use your birth certificate, passport, and driver's license as authoritative documents to prove that you are who you are. As such, the government owns significant aspects of your identity. Depending on your country of origin, the government may collect information about your employment and tax-paying histories, schooling, right to work, marital status, disabilities, religion, health information, military service, travel outside the country, and more. In this sense, you aren't given much of a choice: the government is the authority for identity worldwide.

The benefits of a central authority (that respects human rights) maintaining identity information become clear when you look at populations unidentified by their governments. Across the globe, 1.5 billion people don't have access to a government identity. Of those,

230 million are children, and 60 million are stateless people from areas embroiled in such political chaos that they don't provide a means for identity documentation. These people have actual identities, of course, but they can't prove who they are in an official capacity. In a talk about the United Nations' ID2020 initiative, Dakota Gruener, the initiative's executive director, describes the public health challenges created when people can't be identified.[1] For example, in some cities, the UN has recorded vaccination rates of 140 percent of the children in the population. How can more than 100 percent of the people be vaccinated? More importantly, how is someone protected from disease when the government doesn't even know the person exists?

In theory, governments collect personal data as a means of management: to grant rights, enforce obligations, offer services, and maintain social order. Unfortunately, in practice, governments have a history leveraging data to benefit some people while excluding others. But before I examine the shadow side of government data ownership, let me look at the other major player in data ownership.

Businesses

Beyond government records, your identity data resides in a growing number of places. Organizations like health facilities, financial institutions, e-commerce sites, and frequently used airlines all collect and store data about you. Fundamentally, these entities collect your data to achieve one goal: to make money.

To this end, they collect a variety of types of information, depending on the business. For example, financial institutions collect specific information about their customers and provide specific data about their customers, such as a credit rating. Social networks make money by leveraging the identity data their users share to sell advertisements. Businesses use their customers' data to increase profits by showing them targeted advertising, products, or content; to get referrals to potential customers; to improve services or products; and to optimize consumer experiences and even pricing. The nature and depth of the data businesses collect vary widely, depending on the company. For example, Amazon collects much more data than a company that sells specialty insoles.

Although you may be concerned that the government will use identity data to perpetuate abuses of power, when it comes to businesses, your concerns relate more to the privacy and safety of data stored in their systems and your lack of control over the distribution of this data.

What Could Possibly Go Wrong?

As human beings, we have an innate, proprietary relationship with our identities: I am who I am. My identity belongs to me. But in the digital world, regularly we give and trade away our identity data, putting bits of our identities into other people's control. What happens when we give ownership of our identity to governments and businesses?

Data collection by governments may sound harmless enough to most people, until you consider what happens during war or a change of regime, or when a leader of a democratic country feels threatened.

Throughout history, data collected by governments has been used to profile and discriminate against certain groups of people, to identify and punish opposition parties, to repress activists, and to deport people.

Further, data collection has been used as a political tool. In the United States, political parties use data to draw voting districts to favor their candidates. More subtly, they use data to map the routes of voter registration outreach efforts and to maximize the registration of voters likely to support their party.

A recent *New York Times* article exposed evidence that in Mexico, a democracy since 1917, the government is using surveillance methods to repress journalists, activists, and even politicians who stand up against corruption.[2] Originally this surveillance technology was sold to the Mexican government to deal with drug wars, but today it might be used in sordid ways.

In the past, perhaps you could hide your identity from the government in some situations, but in today's world, it's almost impossible to function—to travel, to open a bank account, to

work—without some form of identity, some of which are more permanent than others. As soon as you have a government-recognized identity and a mobile device associated with that identity, a government or other powerful agency can easily find you.

As biometrics come into play, after you have identified yourself to the government, you can't do anything to erase your connection to that data. You could change your name or pretend to be someone else in a thousand different ways, but your fingerprint and face print are always yours.

In India, the government now identifies people through fingerprints and a retinal scan and has collected biometric data for 1.3 billion people.[3] That's 17 percent of the entire world population. Like most other nations, India has multiple ethnicities and an unfortunate history when it comes to treating groups differently. The vulnerabilities here are clear and growing.

Biometric scanning, consented to by citizens of a nation, is just the tip of the iceberg. Taking people's face prints doesn't technically require their consent, which is exactly what the Chinese government is doing. "A large central database makes it possible for authorities, and some private companies, to identify nearly anyone by capturing their face," reports Josh Chin in The *Wall Street Journal*.[4] With 175 million cameras already installed in China and another 450 million projected, shaming jaywalkers is just the beginning of governments using technology to prevent unwanted activities, whatever "unwanted" may mean.

Once again, the use of this data—or even simply the knowledge that this data is being collected—can produce both prosocial and oppressive results. If there's such a high chance that you'll be identified committing a crime, opposing the government, or even just cutting in line, what are the chances that you'll even attempt those activities?

So the question remains, "How can society make sure every person has a functional identity, while also giving them control over who, including which government authorities, can access it?"

At the 2017 ID2020 conference, Microsoft and Accenture announced *blockchain-based solutions* to resolve the issue of privacy when it comes to biometric information. The biometric data would be stored *off-chain*

so that refugees and other individuals would control who gets access to their identity information. The program is designed to address the more than one billion people who, at the time, had no official record of their own identity.[5] Ultimately such solutions could be the beginning of the return of identity ownership to anyone whose identity is recorded with a government authority.

Overcollecting and Oversharing

The uses of facial printing in China and surveillance technology in Mexico bring to the forefront two additional concerns:

- **Overcollecting:** Although clearly an issue in government practices, overcollecting also runs rampant in business. An online store knows not just how much I paid to buy my child Batman socks but also that I considered Captain America socks. Further, the store recorded exactly how long I stared at the screen and possibly how much I fidgeted with the mouse, suggesting where my eyes were focused before I finally clicked buy.
- **Oversharing:** Not only do entities collect data they may or may not need, often customer data is shared unnecessarily. For example, a radiologist doesn't need to know the name of a patient in order to diagnose X-rays. Similarly, someone granting a loan doesn't need to know the name, race, or gender of the person applying for the loan. Such oversharing can lead to discrimination and bias, which could be prevented if organizations were compelled to collect and share only data deemed relevant.

Knowing what data is being collected, much less to control the sharing of that data, is almost impossible. Even when regulations and agreements are in place about what can be collected, enforcement of these regulations becomes an issue. A recent *New York Times* article reports on US school systems' concerns about Google providing a variety of services to teachers and students.[6] As the lawyers from one school district noted, the legal agreements protecting children's

data were provided by way of a URL from Google. The contents of a webpage can be changed, so an agreement can be altered at any time, leaving no fixed text for reference should a dispute arise.

We as a society would go a long way toward solving the digital identity crisis if we could give people control over the collection and sharing of their data.

Data Sharing: It's Complicated

How many times a week do you find yourself sharing your data? When you make a purchase, sign up for a membership, or go to a healthcare institution, you fill out forms, you provide identifying documents, you slide your credit cards, and you update your health history. In fact, sharing data can be a daily hassle—writing and rewriting your addresses or Social Security number, ticking off little checkboxes, listing your emergency contacts.

Fortunately, these days you fill in fewer manual forms and more digital ones, probably through the Internet, where two main solutions allow you to manage your identity more conveniently. First, built-in browser settings save your most commonly entered information, such as your name, address, passwords, and even credit card information, giving you the option of auto-filling the data when you need it. This data can be stored separately on each of your devices, or you can enable data sharing across devices. Second, you can use a federated identity, enabled through social login, so instead of creating a new identity for every website you sign into, you can sign in using your Facebook or Google account.

Although these solutions simplified the data-sharing experience somewhat, they don't solve the problem of ownership. For example, a federated identity is hosted by a central organization, like Facebook, and your data is owned by that organization, not you. On the other hand, when you enable browser settings to share data across devices, your data is stored on a server of a central organization. Who owns your data then?

What Does Ownership Mean?

Wanting to own your identity makes sense, but in a digital world, what exactly does ownership look like? These sections examine the two distinct elements—your identifier and your data—in greater detail.

Your identifier

Before I talk about owning your identifier, I need to differentiate between identification and authentication. *Identification* is the claiming of an identity, and *authentication* is the act of verifying or proving the claimed identity.[7] Initially, to open a new account or to use a new service, you need to identify yourself. For example, when you create a bank account, the bank representative requires you to show government identification, like a driver's license, to prove who you are. Based on that initial identification, the bank creates its own means for you to authenticate yourself, like an ATM card with a PIN.

For an initial identification, you can use different kinds of identifiers, depending on the service with which you're registering. Some require government-issued identification, like a driver's license or passport. Some register you based on your email address, phone number, or even your social network identity. These identifiers establish who you are; however, although they might represent you, these identifiers don't always belong to you.

Some of the new popular services, like the communication app WhatsApp, utilize your phone number as your identifier. You can also log in to Facebook using your phone number as an identifier. Phone numbers used to belong to phone companies. If you changed phone companies, you had to leave your number behind. Today, in many countries, you can port your number from one carrier to another. In some sense, you own that number, meaning no specific entity can take it away from you.

What about your email address, though? Email began as a means of communication, but your email address has become an identifier as well. If you're using a service such as Google Mail or Yahoo! Mail,

you don't own your email address. When email was simply a means of communication, losing access to your email address would create an inconvenience. But now that your email address has become one of your most important identifiers, if Google Mail or Yahoo! Mail decided to revoke your email address, you could lose access to many of your online accounts.

Have you logged in to a website using your social network username and password? That's another identifier you don't own. If a social media site were to shut down your account, you would lose both your identifier and access to any accounts associated with that identifier.

Your data

Identity ownership extends beyond the identifiers to the data you share online. This data is spread across so many entities that nobody really can know where it is, much less know who has access to it or how it's being used. What would it mean to own your data?

If you generated the data and you don't share it, then ownership is simple: you're the only one who knows it; you're the only one who can decide if you want to share it and with whom. However, much of your data is either generated by others or exists for the sake of being shared. When talking about shared data, the question of ownership becomes more complicated.

I define *ownership of data* as the ability to know the data exists, the ability to control how it is used, the ability to retrieve it, and the right to delete it or share it further. However, although I should be the owner of my credit score, I shouldn't be able to change it. Only the creator of the data should be able to do that.

As things stand now, in the digital realm, all of your shared data is either owned completely by others or it's co-owned by them and you. When others completely own data, either you don't know about it or you can't control it. The data a government has about its people falls into this category, as does the commercial data that certain companies collect and share about their customers. *Co-owned data* is the information you share with certain entities that enable you, at any time, to

view it, control its use, delete, or download it. But these entities co-own the data too, meaning they too can access or delete it.

Some companies are using tricky legal language to maintain their ownership of data. As tech and identity thought leader Doc Searls says, referring to a large ride-sharing company's terms of use, "Interesting legal hack there: you own your data, but you license it to them, on terms that grant you nothing and grant them everything."[8]

Self-Sovereign Identity

Picture a world in which you own your identity entirely. You control both your identifiers and the data associated with your identity. Your identity data isn't held on any institution's servers, and you control access to it, giving (and revoking) permissions to governments and businesses to read information, write information, or otherwise use parts of your identity. This is referred to as *self-sovereign identity (SSI).*

SSI is a concept that evolved within the identity professionals community. In his article, "The Path to Self-Sovereign Identity," Christopher Allen, the foremost thinker and advocate of SSI, has outlined the following ten principles of SSI:[9]

- **Existence:** Users must have an independent existence.
- **Control:** Users must control their identities.
- **Access:** Users must have access to their own data.
- **Transparency:** Systems and algorithms must be transparent.
- **Persistence:** Identities must be long-lived.
- **Portability:** Information and services about identity must be transportable.
- **Interoperability:** Identities should be as widely usable as possible.
- **Consent:** Users must agree to the use of their identity.
- **Minimalization:** Disclosure of claims must be minimized.
- **Protection:** The rights of users must be protected.

These standards are high given the state we're in today. Yet many people would agree on most of these points. Having full control of your data means you get to share what you want, with whom you want to share it, for as long as you want. Allen points out that control doesn't mean that you can change any part of your identity data. For example, you don't want people to be able to alter information about whether or not they have had a particular vaccination. However, you do want to let them control who writes such information in their profile and who is allowed to read it.

Under a system like this, if you shared certain photographs on social media, you could also include them as part of your identity data. If you created connections with particular people through one social network, the fact that you have that relationship could be part of your identity data and could be shared on another social network.

SSI could help businesses comply with privacy-related regulations as well. Such regulations require businesses to provide the ability to export identities. Because each business stores identity in a different structure, transferring identities easily across services would be close to impossible. Under SSI, however, when the customer owns the identity and shares parts of it to businesses, such problems would be significantly reduced.

As the idea of SSI takes hold, there are also a number of attempts to create consortia of entities to manage identity in a consolidated way. The Digital ID & Authentication Council (DIACC) in Canada is a coalition of public and private sectors who is working to create digital identity in Canada. This digital identity would be comprised of attributes that would be verified by governments and the private sector to provide Canadian citizens with a convenient, secure, private, and unified way to identify themselves to both government and commercial entities.[10] Another consortium, made up of ConsenSys, Microsoft, and Blockstack Labs, is working on an open-source, self-sovereign, blockchain-based identity system.[11]

Accuracy, quantity, and expiration

Blockchain technology is promising in allowing people to own all the bits of their identities, but it's going to take some time before that becomes a reality. No matter what technology ultimately enables SSI, people want the ability to control their data in four ways:

- They want to be able to share accurate and timely information.
- They want to expose only the data necessary for a particular transaction.
- They want the shared data to be constantly updated.
- They want the data to expire after it has been used.

Just as in the real world you decide who knows your real name, you also want that ability in your online interactions. For example, if you need to be eighteen years old to make a particular purchase, you should be able to prove that you're older than eighteen without giving any additional information. Furthermore, the information you give to complete a transaction should, if you want it to, disappear from the seller's database at the end of the transaction. If you were in complete control of your identity, in every interaction you would choose what information to share and how long that information could be stored by other parties.

In health settings, financial transactions, and online interactions with commercial entities, when you control the accuracy, quantity, and expiration of your shared data, you can receive the services you need while protecting your identity. Many businesses are aware of the risks involved in managing customer data and of their customers' concerns related to the data they hold. Therefore, they would consider a solution that on the one hand enables them to function, while on the other hand reduces liability and increases customer trust.

How far does SSI need to go to return digital identities to their rightful owners?

The challenge of initially identifying yourself with new online entities is being addressed by solutions such as ThisIsMe, ShoCard, and HYPR, which allow individuals to keep their identification methods on an app that they control. Many of these solutions are based

on blockchain technology, and their marketing messages speak about some aspects of ownership. But they fall short of true self-sovereignty. To use an app such as ThisIsMe, you first need to scan your government ID card, get verification from the government that it's a valid ID, and then use other identifiable information (such as bank validation) before you can make a transaction. These solutions address the technical issue—simplifying the identification process—but the question remains, "Who owns your identity?" "If I need to provide an external proof . . . is this still self-sovereign?" asks Martin Kuppinger.[12]

According to Brennan Wright, ThisIsMe's Head of Marketing, during 2018 the company will release its new distributed digital identity solution, which will utilize multifactor identification technology, eliminating dependence on any one corporation, organization, or nation-state. Such upgrade is taking the app another step closer to providing a true SSI.

Meanwhile, Searls supports a more radical approach to SSI. He argues that a solution can separate identification of the real person from the authorizations he or she has. Entities that interact with you would first get indication about your entitlements and necessary data (attributes), and then, only when absolutely necessary, be informed about who you are. Think of it this way: your real name is irrelevant for many transactions. For example, you have authorization to enter a bar and purchase a beer as long as you're older than a certain age, but your real name doesn't matter in this transaction.

Kevin Hobbs, CEO of blockchain-solutions company Vanbex, agrees with Searls that your actual name isn't needed for confirmation of what you're entitled to. However, he points out, even without your name, you might be identifiable. "Right now via blockchain technology we can map a private key literally to the house and room that you're in, trying to use that private key," Hobbs says, referring to a company called Blockchain Intelligence Group that can provide a granular identification of individuals.[13]

The most mature, comprehensive solution in the market today for SSI is the Sovrin network from the Sovrin Foundation. Sovrin is a public identity network that gives people control over how their personal information is used. Sovrin does this by combining a distributed

ledger, for public discovery of identifiers, with agents that store elements of a person's identity. Trusted entities can create records, called *claims,* that assert information about a person. For example, a person's bank, government, or employer, could create the claims. The person determines with whom to share the information in these claims, and the person can control how much of the information in the claim to expose to any other entity. For example, your employer might assert that you work for her, and you could use this claim to prove you're employed when you apply for a loan at your bank. Over time, the individual manages an identity that includes identifiers, claims, and proofs from others, such as reputation information. Sovrin is set up to allow people to create multiple identities, and for delegation, in cases such as parents managing a child's identity. Although this falls short of complete self-sovereignty, it's still a huge step forward.

Where could SSI go?

"If you own your own identity, you control your own destiny," says Hobbs.[14] Technology-wise, today it's feasible to think of a world in which, when a child is born, the parents could decide not to register that child with a particular country. Over time, each person could amass and control his or her own global, digital identity. If that identity is held not by a government, but by a decentralized system, each person would get to decide what becomes part of his or her identity. Some people might build the kind of reputation that causes certain countries to welcome them as citizens, and therefore they might have more mobility. On the other hand, some people will develop identities that limit their options. In a system such as this, people wouldn't be able to erase records of their activities, but they would have ownership of an authoritative record that expands or limits their options in life.

Reputation and Identity

Today, reputation information is stored in various places. For example, when you sell something on eBay or Amazon marketplaces,

people write reviews of your business activity. Likewise, if you use Upwork or Fiverr, you're rated both as a worker and as an employer. LinkedIn provides reviews of your work, written by people who have interacted with you. If you've stayed at an Airbnb or used Uber or Lyft, your hosts and drivers have rated you (and you've rated your hosts and drivers). Your blog posts and other social media information give insight into your popularity among certain communities as well as your expertise in specific topics. As a matter of fact, even your number of Twitter or Instagram followers has become a type of reputation.

If there were some way of aggregating information like this just in time, everyone could make informed decisions about a person's trustworthiness, qualifications, or experience. Although it's possible today to look up this information on specific websites, an aggregated identity could simplify the process, allowing everyone to evaluate a person's overall expertise, popularity, and skills more quickly than possible, for example, in an interview.

As many people have learned in recent years from studies as well as practice (I recommend reading *Moneyball: The Art of Winning an Unfair Game* by Michael Lewis), when there is access to enough relevant data, assessment based on such data will offer a more accurate picture than will a personal opinion only. Data can bypass biases and help construct evaluations about abstract personal qualities, like reliability and accountability.

And there's more that could be done with such information.

One organization, the Distributed News Network, is looking at reputation and public affairs. Founders Samit Singh and Dondrey Taylor are creating a network where news stories are rated by reviewers as to their reliability, and both the reviewers and writers are developing reputations over time. The system, based on distributed ledger technology, explores the idea that news could become more reliable as a writer builds a reputation. That reputation, amassed from different sources, can verify that a writer is both real and credible. In such a system, readers could rate the reliability of what they see and could request that certain topics are covered, influencing the news to report on topics of interest, in an objective and factual way.

Although the idea remains on paper at this writing, offering factual news based on reputable sources is clearly valuable. Taking the idea further, it could be possible to connect the reputation of a public figure in the news to the articles that are written about him or her. If people rank a news story as reliable, then the facts about a politician in that story could be given a certain credibility rating, which could be added to the politician's reputation as well. Such concept is valid as long as all data, including the ranking, is weighted by its contributor's reputation.

Imagine a future where what people know about a candidate's reputation is more transparent and reliable. This information could assist the electorate in voting. With a reputation checked through such a system, the list of candidates might be better. Although such an election is only speculative at this point, the idea that a person's professional reputation data could be aggregated isn't so far-fetched. Today, looking for a job or a vendor is cumbersome for both the job seeker and the employer. In the same way that people use Yelp or Airbnb today, tomorrow objective measures of someone's history could make this task simpler.

The reputation concept is actually being adopted on a large scale in China,[15] but for troubling reasons. In 2020 a national trust score will become mandatory across China for all its 1.4 billion citizens. This score will represent the "value," or more accurately, the trustworthiness, of each person based on parameters set by the government. In this case, the Chinese government—not the people themselves—will have complete control over its citizens' reputations—who writes to them and who gets to access them.

How Many of You Are There?

Skeptics of SSI suggest that in a world where one person could create and manage multiple identities, identity will become meaningless. However, by using aggregated-reputation data, multiple, substantial, trusted identities can be created.

Of course, you already create multiple personas, to some extent. You have your persona at work and your persona with your friends. Your LinkedIn profile and posts present a different image of you than your Facebook profile and posts do. But even within your closer circles, you may use different personas. You want your father-in-law to know you're an excellent parent to his grandchild, but you might not want him to know you love to dress up on the weekend and carry a lightsaber.

"If you look at society itself, people don't want to be identified as one thing anymore. It's happening in gender; it's happening with race," says Hobbs. "It's nothing sinister; it's just how people feel. We can have two identities, three identities, and it can still be us."[16]

In a self-sovereign world, I can create a new identity for me at any time, but that identity will be meaningless without data. Say that I create a new identity as Lee Chang, and I introduce myself as Lee. That's all fine and good until I look for a job, and I have no employment history. But assume someone gives me a chance, or she knows me from my old identity, and then she gives me a rating as a great employee. I put the money she pays me in a bank account, and I start to get a financial history as Lee Chang. I buy and sell some things on eBay or Amazon, and I grow a reputation as a buyer or seller. I make friends on Facebook and people get to know me. Perhaps commercial or nongovernmental organizations can give me authorizations ascertaining that I'm a safe driver, or a certified professional, or that I'm married to someone. Furthermore, I can earn educational certificates from a variety of sources.

Over time, Lee Chang could become a "substantial" identity and a truly self-sovereign identity. I made it up, I built it over time, and no government dictated what I called myself, what gender I chose, or what permissions I have. Organizations and people could trust this identity because it has credentials and a reputation. Lee Chang has become, in purpose and practice, a real identity.

Delegation

In the self-sovereign world, you may also be able to use your authority by proxy. More than likely you do this in some financial transactions. For instance, I can use my credit card to purchase airline tickets for a relative or child or even give that person a card with a specific amount of money on it. As a matter of course, you delegate people to take all kinds of actions in your name, from the most trivial, such as purchasing a theater ticket, to the most critical, such as deciding whether to keep you on a life-support system if you're incapacitated. You also delegate authority to services or advisors who make investments in your name. You may give a fund the authorization to manage finances within a specific area, for example, specifying a ratio of stocks to bonds and maybe even the stock markets on which you're willing to trade. In this case, the delegation of authority is specific. This person may invest a specific amount of your funds, within a particular set of parameters.

When better technology is available to manage delegations, you'll be able to delegate various kinds of authorization to others. For example, you could authorize voting rights for your stock portfolio, allowing blocks of people to have the kind of influence that today is held by organizational investors.

Delegation capabilities are expanding as technologies, like smart home keys, become a growing trend. Smart home keys allow specific authorization for people to enter your home,[17] enabling, for example, the dog walker to come in once a day at a specific time and on specific dates.

User-Managed Access (UMA) is defined as a protocol "that has been designed to give an individual a unified control point for authorizing who and what can get access to their digital data, content, and services, no matter where all those things live."[18]

As delegation becomes part of your identity in a wide variety of ways, you'll be able to delegate many parts of your lives that, until now, required fairly complex legal arrangements. As self-sovereignty moves closer, you'll be able to give authority directly to a person or organization to do specific things on your behalf, without going through a proxy.

Digital ownership will create a huge shift, not only for the management of your own identity, but for those people who manage their children's identities, delegating permissions to the adults in charge of them at schools, at camps, and during extracurricular activities. Today arcane school district websites and reams of colorful paper manage these authorizations. But self-managed authorization, being simple and more reliable, will ensure that authorization has been done by me, the adult with custody of the child. I have authorized the school to take my child on a field trip or to play sports, or I have authorized another adult to pick my child up from school. That would be a major improvement over the bits of paper I often find crumpled at the bottom of my child's backpack. Similarly, authorization could streamline the process of purchasing prescriptions, managing the care of elderly or ill family members, and handling countless other daily tasks. Such use cases were addressed in a proof of concept done by the New Zealand government in 2015, using UMA.[19]

When Will It Happen?

The idea of SSI has been discussed in one form or another for many years, mainly among the members of digital identity industry consortia. In events that bring together identity enthusiasts from around the world, they explore the question: "What do we need to do to give people control of their identities?" This discussion continues in academic settings, at startups, amongst government officials, consortia and with major tech players like Google and Microsoft.

Technology has arrived to a point where, for the first time, the concept of SSI is feasible. Yet it's still not happening in scale. Why? In the digital age, identity is mainly about connections between people and

organizations. Organizations don't have enough incentive to promote SSI, and the public isn't aware of the significance of SSI. Currently, each unregulated organization determines on its own which identification it accepts, so even if you had an SSI, you wouldn't have places to use it.

Like many other trends, for SSI to take hold, a large organization would need to adopt it, which could bring SSI awareness to the public, therefore raising the demand for similar service from the rest of the organizations as well.

Although SSI is being discussed and shaped, more immediate privacy-related practices are being regulated and adopted across the world. I review them in our next chapter.

Endnotes

1. https://apolitical.co/id2020-john-edge-dakota-gruener-global-digital-identity/
2. https://nyti.ms/2sGmhJ0
3. http://www.businessinsider.com/indian-government-making-a-database-of-citizen-irises-fingerprints-2017-5
4. https://www.wsj.com/articles/the-all-seeing-surveillance-state-feared-in-the-west-is-a-reality-in-china-1498493020
5. http://www.techrepublic.com/article/why-blockchain-could-be-your-next-form-of-id-as-a-world-citizen/
6. https://www.nytimes.com/2017/05/13/technology/google-education-chromebooks-schools.html
7. James M. Stewart, CompTIA Security+™: Review Guide. United States of America: Sybex, 2008.
8. http://blogs.harvard.edu/doc/2016/11/20/ubers-pending-sale-of-your-personal-data/
9. http://www.lifewithalacrity.com/2016/04/the-path-to-self-soverereign-identity.html
10. https://diacc.ca
11. https://cointelegraph.com/news/id2020-how-blockchain-could-be-used-to-solve-global-identity-crisis
12. Martin Kuppinger (founder and principle analyst, KuppingerCole) in a discussion with the author, May 2017.
13. Kevin Hobbs (CEO of Vanbex Group) in a discussion with the author, May 2017.

14. Hobbs in a discussion with the author, May 2017.
15. http://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion
16. Hobbs in a discussion with the author, May 2017.
17. http://www.digital-commute.com/replace-keys-smartphone/
18. https://kantarainitiative.org/confluence/display/uma
19. https://kantarainitiative.org/confluence/display/uma/Case+Study:+Users+Managing+Delegated+Access+to+Online+Government+Services

2

Privacy

Some people don't give much thought to the information collected about them by devices, from cell phones that track their locations, to websites that record their online activities to transactions that identify their whereabouts throughout a day. If you have nothing to hide, why should you care? As long as these devices and functions benefit you in some way, most people are willing to give up some of their privacy. But how much are you willing to give up? What does privacy mean in the digital age?

What Do You Have to Hide?

Cars today are equipped with a variety of systems that track data, and as a matter of course, the companies behind these systems cooperate with police investigations. For example, SiriusXM agrees to provide location information when police serve a valid warrant, and GM has helped in investigations involving recordings from its OnStar telematics system.

In the case of the onboard recordings, consumers have authorized these applications to record their conversations—or have they? Ask Gareth Wilson, who accidentally pushed the emergency button in his OnStar system, which he hadn't even registered to use. Because OnStar is an emergency feature, the audio was monitored in real time, and as luck would have it, Wilson was discussing a possible drug deal.

An OnStar employee shared the recording with the local authorities, who located and searched the car, found marijuana, and served an indictment.[1] When it comes to ethics and privacy, you can look at this story from different perspectives. The customer didn't sign an agreement to have his data shared, but as soon as the OnStar employee overheard the conversation, the employee felt obligated to report it to the authorities.

Of course, most people aren't criminals wanting to hide their misdeeds. Still, no matter where law-abiding citizens live, no matter the benefits of sharing data, most people will say they prefer to keep their private information private.

The idea that some entity—whether it's the government, a certain corporation, or a hacker—can know more and more about my activities disturbs me for two reasons: trust and transparency. Of course I don't want to share my private data with an entity I don't know well enough to trust, because I don't know who within that organization will have access to my data or what they will do with it. Nor do I want to be completely transparent about my life.

When I look at the reasons people choose to hide information, I see a wide range. A person who has committed a crime might want to hide data revealing his whereabouts, whereas someone else merely wants to avoid awkward social situations and choose not to reveal to a sender when he read her text message. People don't want to expose everything about themselves. Think of it this way: Why don't people who have credit cards use them for every purchase? Well, some of them simply don't want *all* purchases to be recorded. They prefer to keep some things—from Saturday night's bar tab to the surprise they bought for their spouse—to themselves.

What Is Data Privacy?

Data privacy isn't about secrecy—it's not about hiding who people are or what they have done. Ultimately data privacy is about control.

People want to share their data with friends, family, healthcare providers, governments, and consumer brands. However, they also want to control aspects of that data sharing: what's shared, who has access to it, and for how long. Everyone understands the need for companies to collect certain information, for example, your age for entering a bar or your shoe size when you purchase a pair of boots. Consumers are willing to share that data, but they also want to have some control over how long that data might stay with a company or what they use the data for. I discussed this topic in Chapter 1 from the perspective of ownership, but as long as you don't really own your data, let me explain what these concerns are and how they can be addressed.

Privacy regulations tend to focus on how an organization stores, secures, uses, and enables a person to view and control his or her data. Regulations are designed to allow everyone to make decisions about how their data is used. By May 2018, organizations serving residents of the European Union must comply with a far-reaching set of regulations, the General Data Protection Regulation (GDPR), which seeks to allow citizens to control the data collected and stored about them. In Australia, legislation seeks to criminalize data *re-identification,* that is, the practice of matching individuals to anonymized data. But some regulations cut both ways. For example, in Russia, a 2015 regulation requires that data about Russian citizens needs to be stored within the boundaries of Russia.[2] This law could be regarded as a power play by the government to ensure Russian jurisdiction and control. Or, as Russian propaganda positions it, this law ensures that Russian citizens have some form of protection when dealing with international corporations or other entities.

As I'll discuss in this chapter, the regulations and tools available today focus on protecting people's data from misuse by organizations. These protections are helpful, but will they actually limit data sharing? As digital artist Daniel Landau says about sharing identity data, people will share even more, "Can it be hacked? Yes. Will it save your life? Yes. Will you allow using it? Yes."[3] In other words, as companies offer consumers ever greater benefits in exchange for their data,

people will continue sharing more and more about themselves. How far will this data sharing go?

Recording Your Private Lives

Virtual assistants, such as Google's Home, Amazon's Alexa, and Apple's HomePod, provide conveniences many people enjoy. You don't need Amazon Dash and you definitely don't need to open your laptop if you can just say, "Alexa, order more laundry detergent." You don't need to pick up your mobile phone if you simply can ask, "How is the traffic on the way to work today?" Yet as you're using Alexa to perform these functions, Amazon is collecting a tremendous amount of data about you, like how often you purchase that laundry detergent, how you pay for it, and where it's delivered. Ultimately, these voice-activation systems will connect with a myriad of other services as well. For example, recently iRobot introduced technology that not only maps your home, but also maps the places that are dirtiest and allows you to give voice commands to your Roomba through Alexa, potentially giving Amazon access to that information. What could be done with all that data? Who knows what a vacuum cleaner could learn about you by collecting your dirt? Maybe those dust bunnies wouldn't reveal much, but what other dirt might your devices dig up—and share—about you?

Some information you might want Amazon to access. For instance, you might want Amazon to predict when you next need laundry detergent. Potentially, specific purchase patterns could indicate the number of people in your household and their ages—data collection you might not want to authorize. Likewise, when you drive, your intention is to have Google navigate for you, but the amount of data stored about your location could have wide-ranging uses, from tracking precisely how many hours you were at work, to revealing whether you're interviewing with your employer's competitor. Clearly, by collecting all this data, your devices come to know you well, and as their capabilities develop, they'll know even more about you.

With today's technology, most of the time virtual assistants don't record data until they hear a certain command, such as "Okay, Google." This explicit voice command provides reassurance that these devices maintain a certain level of privacy. Even with this privacy measure in place, the devices know a whole lot of information that can be used to customize a user's experience. The voice-activated systems know your taste in entertainment, what times you leave and arrive home, when you go to sleep, and when you wake up. As in-home controllers become connected to other devices, they'll know the temperature in your home, when you opened the refrigerator door and for how long, and even how long someone has been in the bathroom.

This kind of data collection has proven helpful in the field of elder-care. Monitoring systems in the homes of the elderly and people with dementia can identify the movement of someone at all times. Using 3D camera technology, several companies are developing algorithms to detect whether a person has fallen and to track his or her movement throughout the space.[4] Although this level of surveillance might seem creepy for everyday use, for the aging population it represents a breakthrough—they can age at home safely, rather than having to move into a nursing home or a skilled-care facility.

The nature of this data is extremely personal, and it's meant to be shared with caregivers or loved ones. The design of these systems requires careful thought about how to share data, so that you get alerts if there is a problem, but you don't view detailed information about midnight trips to the loo. As such, healthcare-home-monitoring systems are designed to learn the regular patterns of the people being monitored and to report only deviations from that behavior. Of course, all the data is stored somewhere with a trusted entity.

As homes become more connected, devices like Google Home, Alexa, or HomePod, which make an easy-to-use interface to anything digital, may also be privy to this kind of data. At such point, all this data will be aggregated within a single service, making it smarter and better.

In-home devices today are used for specific needs, and the user decides when to activate these devices. This opens an interesting question about meaningful consent. As privacy expert Maria Macocinschi

asks, "Could 'okay Google' be regarded as a consent to the processing of personal data?"[5] In the future they could collect and use data in many different ways. For example, any device that plays music could detect if you increase the volume consistently over time. The device, identifying this gradual change as a sign of hearing loss, could notify you, or perhaps someone else, of your condition. Would you be willing to activate a feature like this? In cases where your in-home system notices information such as hearing or mobility deterioration or even changes in eating habits the data could even be sent to a doctor to show the pattern over time. Would you want that?

In fact, voice technology has advanced to the level where it's possible for home systems to identify who is speaking. This data enables you to set your system to identify your children's voices and prevent them from making certain purchases or viewing certain content. Most consumers would welcome such safety controls in their homes in exchange for having a device store their family's voiceprints.

Taking that one step further, what happens to personalization and service when systems identify not only who is talking, but what they are saying all the time(!)? Could a voice system hear a person say, "Honey, where did you put the cookies?" and respond, "The children ate them yesterday"? Of course it could. But would you really want an in-home device recording your conversations at all times? As any parent knows, occasionally we parents say things we'd rather not have recorded and potentially used as fodder for our kids' future appointments with their shrinks.

Just as a music player could detect hearing loss over time, by understanding speech an in-home system could collect useful data. For example, what if voice-identification systems overheard several people in the same neighborhood complaining about the water tasting wrong. One person might think it was just a fluke and not report it, but using collective data, a pervasive problem could be identified much faster. This concept might sound a bit too much like Big Brother watching you, but privacy issues could be managed if the data recorded isn't identified specifically with one person. Would the benefits of detecting toxins in drinking water outweigh your concerns about collecting such data?

Imagine the possibilities for increased public and personal safety. What happens when systems can identify the types of conversations that lead to violent actions? Could these systems have the right to notify the police in cases of extreme violence? Many victims of domestic violence are too afraid to report their family members. Could advocacy groups emerge in favor of using such recordings proactively? Most people would consider this a privacy violation. In theory, what goes on in your own homes is considered your personal business. However, in March 2017, Wikileaks published documents revealing that a CIA surveillance program was targeting everyday electronic gadgets, including smart TVs, smartphones, and even cars, snooping on unsuspecting Americans by turning their gadgets into recording devices. Spy agencies have been able to tap conversations for decades, but in the digital age, the means for doing so are growing exponentially.

As these technologies develop, the fundamental privacy question—are the benefits worth the risks?—will become increasingly more complicated.

The Mega Players

With so much customer identity data held by huge companies, it's no wonder that consumers are worried about privacy and security. When it comes to companies at the scale of Google, Apple, Facebook, Microsoft, and Amazon, privacy concerns rise to a new level. The "Frightful Five," as Farhad Manjoo called them in *The New York Times*, hold an unprecedented amount of information about customers, and it grows by the second.[6]

In addition to their fears about security breaches, people are concerned that their data will be sold or used for something other than its original purpose. Of course, some companies, out for the quick gain, don't handle their customers' data with care. Satisfied by what they can get with a one-off, short-term customer interaction, they're not concerned with developing ongoing relationships.

But the enterprises I interact with see the benefit of creating long-term, trusted relationships with their customers. They are using

people's data for what they say they use it for: to provide more accurately targeted marketing, products, services, and offers to their customers. These businesses have no hidden agendas; they just want to sell more stuff. In fact, most of the business leaders I come in contact with are just like you and me: they want to make money, yes, but they also want to add value to the world. As such, by and large, for-profit companies want to protect their customers' privacy, if not for altruistic reasons, then because they know that thoughtfully collected consumer information is part of their competitive advantage.

Even though the US government's requests for subscriber data have increased, Google has honored fewer of these requests. Presumably, the government makes these requests in the name of law enforcement, but Google's legal team is apparently finding a smaller percentage of these requests to be legitimate. Interestingly, the US government doesn't report these requests, but Google does, offering transparency to its users.[7]

Although many people think of corporations as being locked in with the government, lobbying for legislation that favors their profit margins, when it comes to big tech companies, a trend is appearing: the "Frightful Five" are providing a privacy buffer between civilians and governments. I wouldn't be surprised if at some point in the future, consumers will use their identities with one of the Big Five as their main identity with many organizations accepting that identity in the same way they do government-issued identifiers.

You Aren't as Anonymous as You Think

According to Internet entrepreneur and investor Alexander Muse, the most famous anonymous person in the tech world, the creator of Bitcoin, may have been identified:

"Satoshi Nakamoto gave investigators the only tool they needed to find him—his own words. Using *stylometry,* a person is able to compare texts to determine authorship of a particular

work. Throughout the years Nakamoto wrote thousands of posts and emails, most of which are publicly available. The NSA was able to use the writer invariant method of stylometry to compare Nakamoto's known writings with trillions of writing samples from people across the globe."[8]

This raises the question, "How anonymous is anonymous data?" Aggregated personal data plays an important role in medical studies and research. But before health organizations release medical records, to protect consumers' privacy, they "anonymize" the data, stripping names or encoding phone numbers. Then insurance companies or research institutions can purchase these medical records in "anonymized" form. But are they really anonymous?

In a recent study, researchers found that by cross-referencing anonymized hospital records with police data, they could attribute private medical records to specific individuals.[9] Even more recently, researchers were able to connect personal profiles to anonymized web-browsing data in more than 70 percent of cases.[10] Despite efforts to protect consumers, it appears anonymizing data isn't enough.

"Releasing the data and just removing the names does nothing for privacy," security researcher Vitaly Shmatikov told SecurityFocus. "If you know their name and a few records, then you can identify that person in the other (private) database."[11] Known for de-anonymizing data from Netflix,[12] Shmatikov was featured in *Wired* in 2016 for creating a system that can perform face recognition on a pixelized image designed to hide a person's identity.[13] His team was able to defeat three privacy protection technologies, starting with YouTube's proprietary blur tool.

Regulators are trying to deal with re-anonymization, with the most severe laws being recommended in New Zealand, levying fines of up to NZ$1 million ($700,000 USD).[14] Meanwhile, though they may be using the best anonymizing practices currently available, many organizations are exposing data inadvertently.

Control

People want to share their data, under the right circumstances, but they also want to control what they share and when. According to GDPR, full privacy control would include the following:[15]

- The right to be informed
- The right to access
- The right to rectification
- The right to erasure
- The right to restrict processing
- The right to data portability
- The right to object
- The rights related to automated decision making and profiling

Society is a long way from experiencing full control over everyone's data, but these capabilities are growing. A comprehensive data consent and control solution will need to address the combination of user experience, business goals, and legal requirements.

The evolution of consent

As the Internet evolved, companies have collected consumer data indiscriminately. Through cookies, companies collect users' browsing data with few limitations on how they could use that data. At first, users didn't even know the data was being collected. Now most users are accustomed to getting personalized advertisements informed by their search for a certain pair of shoes or a certain travel destination.

In recent years, there has been pushback on data being shared by *third parties.* That is, if you purchase a flight on an airline's website, you don't want the airline to share that information with a third-party company that would share it further with hotels or other companies,

without your consent. As a result, currently fewer browsers support this kind of third-party data sharing, with some shutting it off by default. As of this writing, Google Chrome still allows third-party data sharing, but in time, in response to user demand, privacy experts predict that it too will change its practices, offering more user-protective defaults.

Further changes are coming into play as well. Many websites, in particular those based in more regulated countries, are now explicitly asking permission to collect personal data, and in many cases, according to their Terms of Use, they won't share that data with other sites. Increasingly, consumer demand and the regulatory environment, with Europe leading, are making first-party data collection the standard. These changes are just the beginning.

How consent works

When a person consents to give data to an organization, he clicks on an agreement that contains a multitude of different clauses, many times without reading or understanding them. Often, this is a company's first handshake with a customer. Some companies care to address this early impression and are getting creative about presenting their Terms of Use in customer-friendly ways. Just as some airlines use humor to engage passengers while presenting their on-board safety regulations, firms like Belgian media company RTBF have turned their Terms of Service into entertaining videos featuring popular celebrities, enriching customer engagement and their credibility.

Another approach to that first handshake is being led by the Kantara Initiative and Customer Commons. These organizations are working together on a User Submitted Terms project to curate icons linked to "user friendly legal terms," describing what those terms mean in common, everyday language.[16] Colin Wallis, Kantara's executive director, explains, "With all the best of intentions and best technology in the world, a good idea can die on the vine if the user experience is flat, unexciting, and non-intuitive. Without a good user experience the capability to control one's own data simply won't be adopted. It's our job as industry consortia to deliver easy to deploy tools, developed by

and for our service provider members, so that they can better serve their customers and the global identity ecosystem at large."[17]

As someone installs a new app, he is informed of all the data he needs to share with the originating company in exchange for trying out the app. It's not possible to check or uncheck part of that data; it's all or nothing. If the person decides later to uninstall that app, he can't rescind his data; it remains with the company. In a world where anyone can create an app and 77 percent of apps are never used after the first three days,[18] it's clear why regulators are pushing for consumer control of this data.

The European Union's GDPR represents government's best attempts to make sure consumer data is used responsibly. GDPR is far-reaching in two ways. First of all, it requires companies to get specific consent for each type or category of data collected. Secondly, it affects any entity that has any data about any EU resident or citizen. For example, if an EU national is purchasing a product on an e-commerce website hosted in the US and owned by an American company, the law applies to this transaction.

In many ways, GDPR represents the ideal for consumers in the following ways:

- If a consumer refuses to provide data that isn't essential to the service, the consumer can't be excluded from using the service. The consumer can go ahead and try out that app while keeping her nonessential personal data private.
- A consumer gets to consent to each type of data collected.
- If the consumer wants personal data removed from a database, the company must oblige.
- A consumer must have the capability to download all the information a company has collected.
- Each piece of data can be used only for the purpose consented to by the customer.

As regulations come into play, many uncertainties remain. Only time will tell how the regulations will be interpreted and enforced and what new practices will arise along the way.

For companies that make money primarily from advertising, abiding by GDPR could be a challenge. Martin Kuppinger of KuppingerCole, says, "If people have a list of the things they consent to, they will say, 'This is the thing that is positive to me. This is the thing that is only positive to Google.' And many of the things that are only positive to Google are the ones where Google earns the money. That might become the challenging part. People ask: 'Why should I be interested in you selling advertising based on my data? I'm only interested in you providing service to me.'"[19] To justify the collection of certain data, Google, as well as many other companies, will need to better align the data it collects with the services the company provides.

In addition, under GDPR consumers can opt out of *profiling*, which means an organization can't assume you belong to a particular category of people just because you behave like people in such category.

The intention of GDPR is to make sure consumers are fully protected. It's not clear how easy these new rules will be to execute and enforce, but what is known is that organizations that fail to comply will face fines of €20 million, or 4 percent of their global turnover, whichever is greater. However complicated the execution of GDPR may be, the spirit of the regulation clearly advances consumers' interests. Additionally, this kind of regulation allows companies to build trustful relationship with their customers while providing excellent services and products for their customers.

Outside the jurisdiction of GDPR, however, consent remains a contentious issue. As I was writing this book, the United States was embroiled in a heated debate over the repeal of Federal Communication Commission (FCC) privacy protections that prevent cable and Internet providers from selling information about their subscribers to third parties. Although the FCC decided to overturn net neutrality,[20] the decision is being appealed. The tone and tenor of this debate showed how deeply people are concerned about their privacy and how strongly they feel that third parties shouldn't use their information.

Some experts see this repeal as a huge step backward, while others say it was more of a debate between different authorities within the government, the FCC, and the Federal Trade Commission (FTC). Although the FCC has authority over communications, the FTC has authority over all consumer activity. Either way, the repeal of these protections has immediate consequences until new regulations are put in place. Given President Donald Trump's insistence that each new regulation be approved only after two others are repealed,[21] consumers could be in for a long wait.

The business end of consent

To serve their customers and grow their customer bases, businesses must keep consumers satisfied. In a digital world, customer satisfaction includes respecting private data. However, complying with consumer data regulations isn't always an easy task, as Flybe learned. According to a BBC article published in March 2017:[22]

> The airline Flybe has been fined £70,000 for sending more than 3.3 million marketing emails to people who had opted out of receiving them. The emails, sent in August 2016, advised people to amend out-of-date personal information and update their marketing preferences. They also gave people the chance to enter a prize draw. But the regulator said Flybe should have obtained people's consent before sending the emails. "Sending emails to determine whether people want to receive marketing, without the right consent, is still marketing, and it is against the law," said Steve Eckersley, head of enforcement at the Information Commissioner's Office. "In Flybe's case, the company deliberately contacted people who had already opted out of emails from them." Flybe told the BBC it wanted to "sincerely apologize" to affected customers. "We can confirm that appropriate mechanisms have already been actioned to ensure that such a situation does not happen again."

The same article reported that Honda found itself in a similar predicament:

> The ICO has also fined carmaker Honda Motor Europe £13,000 after a separate investigation found similar breaches. The company sent 289,790 emails to clarify customers' choices for receiving marketing, but did not secure their consent. "The firm believed the emails were not classed as marketing but instead were customer service emails to help the company comply with data protection law," the ICO said in a statement. "Honda couldn't provide evidence that the customers had ever given consent to receive this type of email, which is a breach of privacy and electronic communication regulations." Honda said it was disappointed with the decision and that it had acted with "the best data protection practices in mind." It added: "It is also important to highlight that we have already taken steps to address the concerns that the ICO has raised, and we are pleased that the ICO has recognised that any breach of the PECR by Honda was not deliberate nor intentional."

Though the fine levied against Honda was minimal, this is merely a preview of what's to come with GDPR when companies will face steep fines even for unintentional breaches.

In order to comply with consent regulations and to protect consumers' privacy, organizations need to perform three functions: collect consent from consumers, log consumers' histories of consent, and comply with consumers' consent. In other words, organizations need to know not only what data they hold in their databases, but also what permissions are associated with that data and how that data is actually used within the organization. As customer bases grow, as marketing technologies develop, as new regulations come online, and as organizations change and expand over time, managing consent becomes extremely complicated.

Think of it this way: when a consumer enters data and consents to its use, the terms and conditions that the user agreed to are attached to that data set. This acknowledgement of consent is for legal use, of

course, to protect the company. It also enables the company to manage the sync between newly launched services and the terms to which users agreed. For example, a website may have collected only an email address three years ago when it didn't have the technology to suggest related products based on user behavior. The Terms of Use that the customer originally signed didn't cover storing behavioral data, so the website releases an updated version of their Terms of Use to enable product recommendations. Now the company will need either to get users to agree to the updated terms or, for users who don't agree to the updated terms, the company needs to provide a different experience according to the terms users had originally signed.

Managing consent to a site's Terms of Use and Privacy Policy becomes further complicated because some term updates require re-consent, and some don't. Some of the updates impact the permissions, and some are insignificant. To protect itself legally, a company needs not only to collect consent, but to prove that its customers re-consented to every significant change in their Terms of Use.

As Flybe and Honda—and many other companies—have learned, a company may think it is compliant, but if a court rules differently, the company may find itself unable to produce documentation to prove compliance, because the company hasn't logged its users' history of consent.

But the complications don't end there. Enforcing consent preferences is a challenge unto itself. Say a customer visits a website for the first time and opts in to a newsletter, using an email account. Such consent would be stored with the email service provider (ESP) the website uses. A few weeks later, say that same customer revisits the website and this time creates an account on the website. This account and its related consents would be stored in the website's database. Now, if the consumer opts out of communication through her website account, she might still get newsletters through her initial newsletter opt-in. Companies that end up in court are required to sync consent data across systems and to show a log of all their customers' consent activities over time and across all properties and databases. For businesses that run multiple properties and utilize multiple vendors, synching consumers' consent becomes a challenge. They turn to companies like

Gigya in search of comprehensive CIAM and consent management systems that can help them collect, manage, enforce, and log consent.

Clearly, consent regulations create significant compliance challenges, but proper consent management serves important protective functions—protecting both consumer privacy and the businesses themselves should they end up in court. Businesses that wish to transform into a solid consent compliance need to go through the following steps:

- **Know which data is stored.** Customers' data is stored in multiple places, both within a company and with the companies' vendors.
- **Understand what data is critical for any given service.** Under some regulations, companies are required to collect only what they need to provide their services, so companies need to identify exactly what they must have.
- **Adjust the company's data collection and consent structure and reword consent requests appropriately.** This is the process of aligning consent to data and the permissions around it.
- **Manage consent when changes are made.** Terms of Use and Privacy Policy changes are common. Companies need to manage customers' consent to these changes, for example, when a website's Terms of Use is updated, prompt returning users to consent again.
- **Allow users to manage the details of their consent.** Companies need to offer solutions for detailed consent management.
- **Secure consent information.** After a company collects consent information, the company must protect it so no one even within the organization can tamper with it.
- **Keep reviewable consent records.** If legal challenges arise, organizations will need to confirm customers' consent statuses at any point in time.

Vendors in the market, including Gigya, the company I co-founded, offer tools to address the challenges of consent and preference management. With such systems on board, customers can see what data a company holds about them, how that data was acquired, and what permissions they granted for data use. Such systems give customers granular control over their data. For example, although it's still a common practice to offer only one type of unsubscribe option for mailings of any kind, advanced businesses are using preference management technology to enable customers to control what types of communications they wish to get, how often they want them, at which time they want to receive them, and what topics are of interest to them. Customers can also easily control which communication channels suit them best for each type of information. A coupon that is valid for today only should probably be sent as a text message, not an email. In this win-win process of fine-tuning, the consumer gets precisely the information she wants, at the intervals she wants, and the business maintains better communication with each consumer.

The Gigya solution helps organizations meet GDPR requirements as well as customer privacy expectations with the following:

- **Preference and consent capture:** Organizations can automate the presentation and recording of consent to agreements for terms of service, privacy policies, cookies, marketing communications, and custom activities.
- **Version control:** Up-to-date records of consent can be maintained for all customers, with tracking of consent history and automated triggering of consent renewals when required.
- **Enforcement of consent:** Preference records can be synchronized with downstream marketing, sales, and services applications, so that organizations stay compliant as they interact across all brands and channels.
- **Self-service preference center:** Customers are given easy access to view, change, export, or remove their information, including personal data, consent to agreements, and communication preferences and frequency (such as subscriptions to monthly newsletters or weekly special offers).

- **A secure data vault:** Customer consent and preference data is stored securely in a cloud-based vault, where it's always available for regulatory reviews, such as the data protection impact assessments (DPIAs) mandated under GDPR.

An interesting concept that could assist in consent management is in discussion within global identity consortium Kantara Initiative. Kantara produced a specification and code for a Consent Receipt.[23] The Concept of Consent Receipt would enable customers to hold a proof in regards to what they agreed to, in the same way a purchase receipt provides a proof that they bought something. Such mechanism would be helpful for the consumer as well as other entities like the authorities and the entity that received the consent.

Who cares?

My expertise lies in helping companies handle their customers' identity data. Over the years, I've observed that most organizations are extremely respectful of both their customers and the data they are storing. Of course, as commercial organizations, they want to make a profit, but ultimately most of them understand that they increase their profits not by manipulating personal data, but by providing better service. In cases where companies have breached user trust, the backlash has been so severe that most companies take great care with consumer data. On the flip side, most consumers understand the interests of the companies with whom they share their data, and they gladly share data in return for better service.

Still many people are overwhelmed by the amount of personal data that they can and do share with companies. Consumers tend to take one of two approaches to this overwhelm: either they share very little, store things carefully, browse the web anonymously, and take other precautions to secure their identity data, or they ignore the problem. The people who chose to ignore the problem may have some privacy concerns, but they feel helpless to do anything to protect themselves. "Privacy is dead," they'll say with a shrug, and they'll carry on with no change in their behavior. This is a normal human response to overwhelm.

Consumer protection agencies and activist organizations are trying to combat this trend, but doing so is difficult given both consumer apathy and the usefulness of the data itself. As long as data sharing benefits consumers, they seem simply to accept that their identity data is held by hundreds of organizations outside of their control. Debates arise around government access to personal data, but the major legal battles generally focus on sequestering data for court cases. People's opinions vary when it comes to law enforcement's use of personal data, but this remains more a political question than a deep personal concern for consumers.

When it comes to companies storing customer data, regulations such as GDPR and solutions such as Gigya assure that data is managed responsibly. These regulations address the data that individuals share with companies, but what about the data individuals choose to share with the people they know? How should that be managed?

Sharing Data with the People You Know

I am often asked: as this trend of personal data sharing grows, will people get more paranoid, or will they simply give up control? My answer is to observe millennials. They demonstrate an interesting dichotomy.

On one hand, they are sharing more information than ever before. On the other hand, at least in the realm of social media and communication apps, they're becoming much more sophisticated about how and with whom they share it. Millennials care less about sharing with organizations while being super sensitive about the data they share with people they know. Members of the younger generation are both highly active on social media and quite sophisticated about using the available tools to manage the data they share.

Snapchat, Facebook, WhatsApp, Instagram, and other new communication tools understand this trend and get into the nitty-gritty of privacy: person-to-person privacy controls. When a person messages

a group in WhatsApp, she can track not only who saw the message, but also exactly when each group member saw it. In some apps, you can hide from others whether or not you viewed their messages. You can now also send messages that expire after a limited time. Snapchat seems to be the leader in offering new privacy-related features and controls. Not only is Snapchat the first to come up with the concept of a message that expires, but the app also offers additional privacy capabilities, like the one enabling a sender to know not only if the recipient viewed the video, but whether she replayed it. Snapchat goes as far as giving an indication about whether the recipient took a screenshot of the video you sent!

This granular control is important, because people are realizing that seemingly benign features of some apps share data they don't want shared. For example, when an app shows that you sent a text late at night, the app is notifying the recipient you are awake and storing information about your sleeping habits. Is that what you want? Nearly everyone has been in emotionally charged chat conversations where they can see how long it takes for the other person to type a response. Does the sender want that information shared? Similarly, do you want people to know you saw the message and didn't respond?

What other information are people sharing inadvertently? An employer can tell based on someone's status on Skype whether she is at her desk, moving the mouse around, or away from her desk. Is that something you want to reveal? What about the information URL autocomplete exposes to anyone who uses your computer? Have you noticed that previous queries are also visible when you search through Google Chrome's URL address bar? These small information leaks are affecting users' lives and their social identities on a daily basis and in meaningful ways. For the younger generation, this is what privacy means.

You're probably familiar with a different kind of information leak as well—when people in your inner circle use your data inappropriately. In these cases, freely shared data, for example, a personal photo, is forwarded or shared in a way that the originator feels is a violation of her privacy. Each person's value system is different; therefore, when you share data, you also trust someone else's judgment. Of course,

the problem of shared personal data has always existed, as you know if you've ever told a secret. In a world where so much data is shared, the risk of exposure, hurt feelings, or unintentional damage rises. As more data is stored online, it becomes even easier to share, either intentionally or unintentionally. Your most intimate correspondence is now available, and something as simple as a screenshot can be used to share widely a conversation that was intended for just one person.

Young people understand this profoundly. They care about their privacy, and that's why applications such as Snapchat are popular with their generation. These advanced communication apps can expose a lot of information, but they give more certainty about how data is shared. If you ask teenagers about their groups in these apps, you'll often get answers like, "This is the class chat with the teacher in it, and this is the class chat without the teacher, and this is the one where we just talk about the parties going on." They have a myriad of sharing circles, and they care what data goes to whom. In fact, they care so much that many teens create real and fake Instagram accounts, hashtagged `#rinsta` and `#finsta`, posting different content to different audiences on each.[24]

When you share personal information, you do so trusting that employers aren't going to scold you for being away from your desks or that your partner isn't going to flip out if he sees your status message is still alive late at night. But in some cases, that information can cause serious consequences, whether it's with a jealous romantic partner, a micromanaging boss, or an overly involved parent. As I have discussed, regulations and data management solutions play important roles in business operations. But they don't solve the issue of data misuse perpetrated by people you know. Ongoing discussions about data privacy ultimately miss this point: How do people manage data shared with the people in their lives?

Today's trends tell you that customers, especially the younger audience, care a lot about privacy, but it's a different type of privacy. It's the person-to-person privacy. In fact, privacy is the reason behind the birth of Snapchat, a product that brought the extraordinary market value of more than $20 billion. These privacy practices represent a new way of controlling data, and the most advanced customer-facing businesses

are taking note. Members of the next generation want granular control over their data, yes, and when you give them control, they're likely to share more.

Regulations, consent solutions, and granular privacy controls help consumers feel more comfortable sharing their information with commercial entities and other organizations, but they don't address another major consumer concern: data security. I take a look at that in the next chapter.

Endnotes

1. https://www.forbes.com/sites/thomasbrewster/2017/01/15/police-spying-on-car-conversations-location-siriusxm-gm-chevrolet-toyota-privacy/#42c745372ef8
2. http://www.mondaq.com/russianfederation/x/435890/Data+Protection+Privacy/Russias+Personal+Data+Localization+Law+Goes+Into+Effect
3. Daniel Landau (digital artist, researcher, and lecturer) in a discussion with the author, January 2017.
4. https://www.researchgate.net/publication/229138750_Home_monitoring_of_elderly_people_with_3D_camera_technology
5. Maria Macocinschi (privacy expert, doctoral candidate, University of Turku), in a discussion with the author, December 2017.
6. https://nyti.ms/2q31gas
7. https://www.google.com/transparencyreport/userdatarequests
8. https://medium.com/@amuse/how-the-nsa-caught-satoshi-nakamoto-868affcef595
9. http://techscience.org/a/2015092903
10. https://thestack.com/security/2017/02/07/72-of-anonymous-browsing-history-can-be-attached-to-the-real-user/
11. http://www.securityfocus.com/news/11497
12. https://arxiv.org/abs/cs/0610105
13. https://www.wired.com/2016/09/machine-learning-can-identify-pixelated-faces-researchers-show/
14. http://www.zdnet.com/article/nz-privacy-commissioner-recommends-australias-data-re-identification-criminalisation-lead/
15. https://gdpr-info.eu/
16. https://kantarainitiative.org/confluence/display/infosharing/User+Submitted+Terms+project+overview
17. Colin Wallis (executive director at the Kantara Initiative) in a discussion with the author, December 2017.

18. http://www.androidauthority.com/77-percent-users-dont-use-an-app-after-three-days-678107/
19. Martin Kuppinger (founder and principle analyst, KuppingerCole) in a discussion with the author, May 2017.
20. https://www.fcc.gov/restoring-internet-freedom
21. https://www.theatlantic.com/business/archive/2017/01/trumps-regulation-eo/515007/
22. http://www.bbc.com/news/technology-39430349
23. https://kantarainitiative.org/confluence/display/infosharing/Consent+Receipt+Specification
24. http://www.howcoolbrandsstayhot.com/2017/in a discussion with the author/how-teens-get-real-with-a-fake-instagram-account/

3

Protecting Your Identity

While I was writing this book, French Defense Minister Jean-Yves Le Drian made numerous calls to French companies in the defense and security industry, requesting assistance for their country. Using a secured video line, he explained that the country needed funding for a top-secret project, which of course he couldn't disclose, and of course, if the companies agreed to donate funds, they couldn't disclose the nature of their donations to the government. He gave them the details of a secret account, and some companies transferred money to European bank accounts, which, as it turned out, were connected to entities in China.[1]

This may sound like government corruption, but in fact it was an elaborate cybercrime. All the details had been faked—including a man made up to look like Le Drian, speaking from an office furnished like his, apparently located in the Middle East. This false identity had been set up with all the identifying features, such as email and phone exchanges that made the companies think that they were, indeed, speaking to the French Ministry of Defense. The case is troubling for two reasons. First, the minister's identity had been thoroughly falsified, from his digital to his physical identity. The second, and perhaps more shocking part, however, is that these were security companies, and they were duped just like any other company under a cyberattack, losing millions of euros.

As Martin Kuppinger, principal analyst at KuppingerCole, said in his keynote address at the 2017 European Identity and Cloud Conference, "Every individual and every organization is or has been

attacked successfully." According to Assaf Mischari, former CTO of the 8200 intelligence unit cyber division in the Israeli Defence Force (the unit, according to the press, responsible for the Stuxnet virus injected into Iranian nuclear plants), "Almost any cyberattack begins with identity."[2]

In 2013 and 2014, as many as 3 billion identities were compromised through an attack on Yahoo! alone.[3] The stolen data didn't include sensitive financial data, such as credit card or bank account information, but it did include passwords, birth dates, names, and telephone numbers, meaning people's identities were compromised and that data could be used in any number of ways. In June 2015, the US Office of Personnel Management experienced one of the largest breaches of government data to date,[4] exposing the private data of up to 18 million current, former, and potential employees.[5] In September 2017, Equifax announced an identity theft event potentially impacting more than 143 million people. According to the FTC, "The hackers accessed people's names, Social Security numbers, birth dates, addresses, and in some instances, driver's license numbers. They also stole credit card numbers for about 209,000 people and dispute documents with personal identifying information for about 182,000 people. They grabbed personal information of people in the UK and Canada too."[6]

Though they don't always make the headlines, data breaches happen regularly. As everyone's online identities take a growing part in their representation, they become increasingly vulnerable to such cybercrimes. If your identity is allowed to do it, it will be done, and not necessarily by you. Hence it's important to know who has access to your identity data and how you can best protect it.

Who Are You?

Your identity is important to you, obviously, but it's also important to other entities that need to affirm who you are, and to know what privileges you should have, or to *authorize* your access. In some places, like airports, elaborate procedures verify people's identities and privi-

leges. The check-in process at the airport reviews your identification, authenticating your identity, and authorizes your right to travel, affirming that you have purchased a ticket, that the ticket is in your name, that you have the right to enter the country of your destination, that your bags belong to you, and that your bags don't contain any forbidden items or materials.

You can see how authentication and authorization work in other areas of your life, too. If you have a bus or theater ticket, you don't need to authenticate your identity; you just need to provide the authorization, in the form of the ticket. To perform online banking, you need to authenticate with a password or fingerprint. Usually, that's enough to allow you to transfer money between your accounts. But if you want to withdraw a large amount of money, you might be required to go through a different level of authentication. To close or open an account, you may even need to go to your bank in person.

It used to be simple

In real life, you usually don't use any metrics to authenticate people. You know your friends and colleagues immediately by sight, voice, and behavior. In addition, you might see their names or photos on your phone screen when they call.

Before the era of big governments, identification was simple. When you were born, your parents gave you a name, perhaps registered you somewhere, and from that moment on, that was your identity. In some places in the world, that's still all people have. In other places, like India, where 25 percent of births are unregistered, they don't have even that.[7] In fact, with today's mass migration and refugee situations, many of these dislocated people come from communities where no documentation was required. They lived in a village where they knew their neighbors by name and face. Today, in many of these villages, some people may perform banking and other activities using a mobile phone, but for the most part, the main form of identification for these masses of people remains their physical presence. They don't hold or need government identification.

In fact, government identification is a fairly new concept. Iceland began recording births in the eleventh century, because the population was so small they wanted to avoid marriages between relatives. In England local parishes starting in the mid-sixteenth century recorded births, but not until the nineteenth century did the government begin keeping records, and documents such as passports started to become common forms of identification. So when talking about undocumented migrants, it's worth considering that almost all humans were undocumented less than 200 years ago. Until recently, most children under sixteen had no physical documentation to show unless they travelled internationally, and today, many countries don't require citizens to carry a national ID.

In the past, people didn't need to register major life events with the government. Obviously, driver's licenses were required only after cars were invented with the first ones being issued in 1899. In every tradition, marriage was a public ceremony, because the community needed to know that this woman and this man were promised to one another, but no documentation was required. Because identification was a community-authorized activity, the community—not the government—was the forum for official transactions such as marriage. For financial transactions, until recently, most people personally knew their banker. Signatures were used as a form of authentication, but they were needed only by a small minority of people who did business across large distances. Most people stayed in one place or moved very rarely, and their community knew them. Clearly, identity theft wasn't the pervasive concern then that it is today.

A digital identity

Digitized identity, again something taken for granted today, is still in its infancy. With the evolution of digital technologies and more recently the Internet, additional methods of proving who a person is have come into play with passwords to magnetic cards and biometric recognition. Today's methods of authentication are neither simple nor secure enough.

How much time do you spend proving who you are? How often do you enter a password, wait in line to show your identification, sign a credit slip, show a membership card, receive a verification code via text, answer a security question, or even search for your keys? Every moment you spend authenticating is a moment of time wasted, and it's not the fun kind of wasted time, like enjoying an extra five minutes in bed. When you consider that for most of human history it took no time to prove who people are, it's overwhelming just how much of life people spend identifying and authenticating themselves. For a culture that claims "time is money," it's strange that we as a society give so much time away, for little gain. Not only are these authentication methods a major time suck, they're each vulnerable in their own ways, failing to give us the foolproof security we seek.

A dead man walks into a bank . . .

It sounds like the beginning of a joke, but it has happened to thousands of people in India, and it happened once to security and neuroscience expert Moran Cerf, who was contracted by the Israeli government to check that his new citizen registry was secure. On a whim, Cerf's team marked him as "dead," to test the system, thinking it would be just as easy to change back. But when Cerf went to his bank the next week and the bank was unable to serve him due to his demise, he began to learn just how hard it would be to correct that one data point.

When you die, the Israeli government triggers a series of events, including informing your bank and sending a consolation card to your parents, which in turn triggers tombstone vendors to send advertisements to your family. It took Cerf a month to get his life back, but in India, it can take eighteen years,[8] as the more than 10,000 members of the Association of Dead People know. It's not surprising that organizations that store your identity have protocols to follow when you pass away, nor is it surprising that they don't have a protocol to follow in case you come back to life.

Unfortunately, if your data is compromised—whether you're prematurely marked deceased or someone has stolen your Social Security number—it's possible for someone to take over your life in more ways

than one. In the United States, where many people don't have a picture ID, and in India, where many provinces use antiquated identification systems, family members have been known to steal identification to collect pensions or unemployment payments. As Cerf explains, "In the US, there are legal citizens with no ID, and they get a temporary paper to prove their identity for government benefits, like a pension. You're supposed to get a permanent one, but nobody does, and then you're in the system, but there's no picture of you, and the temporary document has no picture. Your brother could steal it and take your pension payments. It's an official document, but it's not identifiable. So there's a whole world of crime between families." Cases abound of people using fake identification of relatives, dead and alive, to receive benefits in the US. [9,10,11,12,13,14]

Identification fraud is a problem both in countries where citizens aren't registered and in countries where identification systems are easily manipulated. In fact, in the US, fraud involving total strangers stealing your identity is more common than having a relative take over your name.

"What would you do if your child were in foreclosure on a home in another state?" a Carnegie Mellon report on child identity asked.[15] In studying 40,000 children's identities in a specific region in the US, researchers found that 10 percent had been used by someone else, compared to only 0.2 percent for adults in the region. (The national average for adults is closer to 1 percent.) Because children's Social Security numbers are "clean" of any financial records, they're easier to use—and to attach any name to—than adults' numbers. As a result, a person can take over another person's identity without knowing anything about that person or having any resemblance whatsoever to the person he or she is pretending to be.

Of course, if you know some information about a person, it's even easier. Knowing someone's name and phone number or email address can be enough to get control of that person's entire life.[16] And that's just on the individual level. On a larger scale, organizations can be hacked—as Yahoo! and Equifax were—exposing large quantities of information. An entire industry of identity theft prevention tools has grown up around this specific danger.

When a person's identity is stolen, two questions come into play: How do you get this person a clean identity? and What do you do with the data attributed to the fake identity? Moran Cerf, who learned firsthand how difficult it is to bring an identity back to life, was contracted by the US President's administration to help answer these questions. As long as identity fraud is possible, we as a society need to consider both how to help secure people's identities and also how to help the people whose identities have been compromised.

Government, security, and identity

As identification is becoming digital, governments are looking for improved methods of identity verification. For a number of years now, anyone entering the United States must give fingerprint identification when presenting a passport at airports and major border crossings.[17] At this writing, sixty countries offer biometric passports, either optionally or exclusively.

Estonia, searching for a competitive edge as a country, has become the leader in this area, allowing its citizens to perform all of their government-facing actions online, using an online identification. It's possible to apply for e-citizenship of Estonia, in fact, without ever setting foot in the country. With Estonia's security protocols and online identities, it's feasible that the entire country could be physically evacuated but none of the government or banking functions would be impacted, because everything is online. Imagine, if all governments take Estonia's lead, how the world will change.

Crime and Identity

Identity data has proven extremely useful in preventing and solving crimes. Already, police are seeing dramatic reductions in hit-and-run accidents,[18] due to the ability to identify people by their phone locations as well as the prevalence of surveillance cameras at intersections. People are less likely to leave the scene of an accident when they know they're likely to be found, and an innocent person is less likely to be

held responsible for an accident when these tools are used to prove he or she isn't at fault.

The 2016 movie *Patriots Day* outlines how authorities investigating the Boston Marathon bombing were able to track the perpetrators through surveillance videos in the vicinity and by backtracking the bombers' steps. In today's world, this is deep investigative work. In tomorrow's world, data about people's whereabouts and activities may be so easy to access that just knowing you can't get away with a crime will be a powerful deterrent to committing one. It's unlikely that the world will be crime free, but it will be much harder to get away with a reported physical crime.

Although offline crime becomes more difficult to get away with, online crime is on the rise. It's now technically possible to take over a hotel and lock its guests in their rooms until a ransom is paid[19] or to take over the steering and braking of cars.[20] Committed remotely, these crimes are difficult to trace back to the perpetrators. As the temptation to commit online crimes increases, so will the scope and scale of such crimes.

In fact, cybercrime has become an industry that functions like other organized industries. Cybercriminals even sell hacking tools over the Internet. Furthermore, cybersecurity expert Hanan Levin describes entire business models similar to revenue sharing in this black market. Here's how it works: first, Hacker A uses software that taps into the computing power or the connectivity of computers that belong to innocent people, then he rents it out through an affiliate model. When a different hacker, Hacker B needs increased computing connectivity, for example, to run software that attempts to hack major sites by trying different passwords, he can rent increased connectivity from Hacker A. For every breach that succeeds using these rented computers, Hacker A gets a specific payout. This affiliate model means that even hackers equipped with the most basic tools can take over your computer, using their limited skills as an income generator. This also explains why even a home computer, with no important data on it, is still a target for hackers—your laptop is a potential source for expanded connectivity or greater computing power.[21]

What's it worth?

When you expose basic personal data, such as your daily habits and locations, you become vulnerable in a variety of ways. Even without knowing your most private data, a person can guess that your house is probably empty because you checked in at a vacation spot or a coffee shop. A person can guess that you're tired this morning, because you were active on social media late last night. A person can learn all this about you without any attempt to access data you were trying to hide. With data you're trying to keep secret, such as your credit card information, the potential for damage is much greater.

Identity data is valuable to cybercriminals for several reasons. People could make purchases with your identity, from buying expensive electronic devices and appliances with your credit card to applying for a loan with your Social Security number.

Other attacks include stealing data and threatening to make it public, or locking people out of their own data. These hacks have been made relatively easy by software like packet sniffers and ransomware, which allow a hacker to hold data while threatening either to make sensitive data—like medical records—public or to block access to data—like the family photos on your home computer—until payment is made. The data on your computer isn't intrinsically valuable to everyone, but to you it's worth hundreds or even thousands of dollars. Additionally, I, and other industry leaders, are seeing a worrying increase in complete account takeovers (ATO), which almost always happen through password authentication.

Although hackers access individuals' data and passwords in a variety of ways, no matter how the data was acquired originally, it's easy enough to get one's hands on it in the Dark Web.[22] (The *Dark Web* refers to websites and pages that are almost impossible to find. It's dark because it can't be found by using traditional search engines or visited by using traditional browsers.) On this black market, different forms of stolen data sell for a range of prices. While in 2014, a medical record was worth several hundred dollars, today the price is down to less than $10, because these records have become so common.[23] The price of credit card data also is surprisingly low.

In addition to the obvious data available, such as passwords, credit card numbers, and health records, a blog post from Symantec showed a myriad of different types of stolen data available on the Dark Web:[24]

- Scans of real passports ($1 to $2), for identity theft
- Stolen gaming accounts ($10 to $15), to yield valuable virtual items
- Custom malware ($12 to $3,500), to divert payments to the attackers
- A thousand followers on social networks ($1 to $12), for marketing
- Stolen cloud accounts ($5 to $8), for hosting a command-and-control (C&C) server
- Sending spam to 1 million verified email addresses ($70 to $150), for marketing and phishing
- Registered and activated Russian mobile phone SIM card ($100), for communications

Whether it's a dollar or $5,000, stealing data pays.

Phishing for data

Some hackers rely on social engineering—manipulating humans to make judgment mistakes—to access data, something that security technologies still haven't addressed well. For example, a targeted phishing attack sends infected emails to employees of a particular company. Even though only a fraction of the employees might click on them, one is enough for the company to become infected. Can companies that offer security solutions develop a solution to protect people and organizations from human error?

Security companies look for different ways to address this problem. One company's efforts ended with a surprising twist. Its story, well known in the security industry, sounds unbelievable, but I've met the company's CEO, and he confirmed its truth. To ward off a phishing attack, this startup created a slew of fake email addresses that eventually overwhelmed the attacker with false responses. This tactic

proved so effective that one day the CEO of the company received a phone call from a Russian Mafioso who said, in short, *Your anti-phishing software is very effective, and it's putting me out of business. I advise you to close down the company. Otherwise, I'll kill you and your family*. The CEO called the FBI, and the FBI concluded that the phone call was a credible threat. Indeed, that Russian man would likely be as good as his word and murder the CEO and his family, so the CEO shut down the company. As this CEO learned, entities in the hacker industry aren't only well organized, they're also rabidly protective of their territory.

Who protects you?

For crimes in the physical world, people call the police who come to the scene of the crime to gather clues, begin investigating, and eventually (hopefully) apprehend the perpetrators. In the case of cybercrimes, the crime scene is *distributed,* which means the perpetrator could be anywhere in the world, the victims could be somewhere else (or in multiple locations), and the servers are in yet another location. When a cyberattack occurs, the authorities may or may not be called in. Often, when a company is breached, the company resolves the problem without reporting it. (However, regulations in many countries do require companies to inform customers if their records have been breached, and companies who suspect a breach must warn their customers to change their passwords.) Only in extremely serious cases will the FBI or Interpol get involved. In other words, when it comes to cybercrime, consider that you aren't protected by the authorities, and ultimately there are only two lines of defense: the security industry and your own personal choices.

A hacker's ROI

As crime moves online, it's also becoming more sophisticated. One person's data may be a fairly easy target, but its value is low on the open market. However, as you collect more data on yourself, its value becomes higher to you; therefore, the potential return becomes higher for hackers. How much would you be willing to pay to rescue all of

your tracked data for the last five years, as well as your photographs and other personal information?

As privacy protection regulations come into place, they will require functionality that enables individuals to download all their data and erase it from the cloud (right to data portability and right to erasure).[25] Ironically, that could make the ransomware hacker's job even easier. In the future, someone could possibly pose as you, download your identity data from the organizations that store that data, and then have it erased from the cloud, holding it for ransom, making sure that you can't get to the data, even if you eventually do regain access to your account.

As businesses engage this technology, they need to get ready to address these security issues. How will they quickly and easily verify that a person who claims their account has been hijacked is indeed the owner of the account, when this person doesn't know the new, hacker-made current password? (Note: again, this is an authentication problem.) How can a business keep backups to easily recover a hacker-deleted account, yet still comply with regulations that require businesses to provide a means for users to delete their data permanently?

Until recently, individual medical accounts were mostly held for ransom, but today ransomware is targeted at the hospitals themselves, because the return is higher.

As you'll see, opportunities to hack both individuals and organizations will continue to grow as hackers develop and share new tools to enable attacks and as companies fail to engage even the most basic security measures.

High volume, low-cost hacking

Professional hackers target databases that store large amounts of data about many people. The number and sophistication of hackers is increasing, but even fairly unsophisticated hackers, known as *script kiddies,* can obtain free tools that exploit some basic vulnerabilities. For example, in one of the most common hacks, software enables a hacker to access any open Wi-Fi network and record keystrokes and entries made on that network. In this way hackers can acquire credit card numbers, passwords, and anything else a user might type while on a

public network, on sites that aren't using secure protocols. Similarly, hackers can put sleeves over the opening of an ATM to record banking data and create fake ATM cards based on stolen data. Many organizations have protected themselves from these basic attacks, but some haven't implemented the proper security, even against an unsophisticated hack.

The security market continually advances to meet these new challenges, but it functions like any ecosystem. As the protection gets better, the attackers get better too. Each level of attack leads to the next level of security, which leads to the next level of attack, and so on. Organizations that invest more on protection stay ahead of the curve and become the more difficult targets, but in the low end and middle of the curve, many organizations remain exposed to basic exploits. Every day, the tools used by malicious hackers become cheaper to buy and easier to execute, which further exposes those organizations with low levels of investment and security standards. Therefore, stagnation means increased vulnerability.

In fact, some of the more serious security problems can occur at the lower end of the curve with organizations that don't see themselves as important targets, because they're smaller businesses or because they don't store credit card or personal data. Although they might not have much to protect, because users duplicate passwords across many platforms, stealing a password list from a low-risk site with low levels of protection may allow hackers to use those passwords on sites that do store that kind of data.

Scaled-up exposé

In the past, a criminal who wanted to steal sensitive data had to crack a safe, remove physical documents, and make copies of them for distribution. The digital medium, however, makes data theft more efficient and effective. After hackers obtain authorization, they get superpowers that enable them to access, copy, and distribute enormous amounts of data with very little effort. Theft of data at that scale simply wasn't possible before.

One of the largest exposures of data in history was perpetrated not by a hacker, but by Edward Snowden, a man who was authorized to

access data in the US government's most secret databases. Snowden felt that the government was wrongfully spying on citizens, and he made tremendous amounts of data public. In the same way, when content at the size of a library could be copied onto a drive the size of coin, breaches of company databases, many times by authorized personnel, have the potential to expose invaluable data at scale.

Authentication

As you can see, many hacks capitalize on authentication methods that are deeply flawed. Today's authentication methodologies are based on three pillars:

- **Something you have:** Something you have would be a device, a key card, or any other physical object.
- **Something you know:** Something you know is a password, an answer to a security question, or any other piece of information.
- **Something you are:** Something you are is any biological attribute like a fingerprint or voiceprint.

Multifactor authentication (MFA), or *Two-factor authentication* (2FA), utilize a combination of two or three of these methods, most commonly a password and a device. This approach is proven to be radically more effective in protecting online accounts. The limitations of current technology make MFA cumbersome for users, but systems are coming into place to improve the experience and reliability of various forms of authentication.

The death of passwords

Passwords are the number one way of protecting yourself online. Yet passwords are poor forms of authentication for a number of reasons. First, people tend to choose passwords that are easy to guess. Secondly, people often use the same password for multiple logins, so if one database with non-encrypted passwords is compromised, the

stolen password can be used as credentials for another, more important login.

Passwords fundamentally pose problems for people. The human brain isn't designed to remember passwords, especially not random ones, so many people choose simple words or patterns with known numbers, such as their phone number or birthday, which are easy for hackers to discover and guess. Most websites protect people from choosing too simple a password by setting up length requirements, necessitating the use of numbers and special characters, and requiring users to change the password periodically. Although doing so seems like a great idea, in fact, it forces people to remember things that their brains generally aren't good at remembering. So what do you do? You write the passwords down somewhere, either on a piece of paper or in a simple text file, exposing yourself to risk.

Why do websites ask you to choose such complicated passwords? Well, obviously you don't want to be like 17 percent of the people who use `123456` as a password,[26] because a hacker tries this combination first. For that matter, don't use any of these passwords, which was the ten most common in 2017 by Splashdata:[27]

- 123456
- Password
- 12345678
- qwerty
- 12345
- 123456789
- letmein
- 1234567
- football
- Iloveyou

But what are the chances that a hacker would guess that someone uses `heyjude` as a password? Isn't that secure enough? There's a good reason for these complex passwords, one rooted in yet another hacking tool the Dark Web offers—lists of hashed strings.

Until a few years ago, many databases stored passwords as is. If the databases were hacked, they'd reveal each username and password in their lists. Today, most passwords are stored as a hashed code, which means, when the hacker gets it, she doesn't get the actual password but a set of characters that are irreversible.

To make that hashed password useful on other sites, a hacker must first convert it using a table that shows the hash of all common character combinations, from common names, to song titles, to millions of number combinations, to hundreds of millions of other combinations users believe to be unique. A table will likely include a common number combination like `654321`, but more complicated passwords, like `eG3#fe74`, are unlikely to be included. Therefore, they're safer although they're also unfortunately impossible to remember.

If a person does use highly secure passwords and different passwords for different sites, those passwords need to be recorded somewhere so the person can remember. Some people use physical lists written on pieces of paper, but more commonly, people keep their credentials in a file, on their computer, or in the cloud. That single file isn't secured, and it can be hacked.

Password managers offer a technological solution for storing all of your passwords under a single—that's right—password. The advantage of using a password manager is that the app itself is secured, unlike a file, and can be used across a variety of devices. However, it's still a single point of failure. "Password managers are society's method of moving bad habits to the computer," says cybersecurity expert Tyler Reguly. "It's bad form to 'write down' passwords, so instead we 'store' them on our computer. 'Store' is simply the digital equivalent to 'write down.'"[28]

Interestingly, the US National Institute of Standards and Technology (NIST) is proposing new guidelines for passwords that don't require cumbersome combinations of letters, numbers, and symbols, nor do they require periodic changes. Instead, when combined with additional factor authentication, the proposal compares new passwords to a list of nonsecure passwords and simply forbids users to choose a nonsecure password. For example, a user won't be allowed to choose the password `qwerty`, nor will she be forced to choose something as complicated as `@34Gr%D`. A simple and unique string like `brainpop` will do.

Figure 3-1: *Does this look familiar?*

This move away from highly secure passwords is enabled by the growing adoption of multifactor identification, which can maintain or even increase the current security level, with less complicated passwords.[29]

Beyond passwords

MFA typically requires people both to know something, like a password, and to have something, like a mobile phone or a special key card. Some sites require MFA, and others, like Google, allow users to choose the level of security they want.

Unfortunately, recent reports have shown that MFA using text messages is no longer reliably secure.[30] In a case in which authentication requires a password and a cell phone to receive a verification code via text, a hacker who has your password can buy a new phone or SIM card, then fool a phone company into redirecting text messages to the new device or card in your name.[31, 32] Therefore the hacker has access to your verification code.

More organizations are beginning to use risk-based authentication (RBA). If you've ever been asked to re-enter the last four digits of your credit card or answer a security question, you've encountered RBA. The site you logged into detected something about your behavior that was outside of your norm. Perhaps you were using someone else's computer or you were logging in to your bank from a new location, so the system required you to confirm your identity. Simple security questions are vulnerable too since their answers are many times available on social media. Forrester Research analyst Fatemeh Khatibloo advises people to delete or un-public Facebook posts that reveal information like the concerts you attended.[33] RBA technologies, however, gain sophistication through a combination of a growing number of signals and advanced algorithms.

Google became a leader in securing online identities. In an attempt to give users more control, Google now provides an account center for each user, allowing individuals to choose their security level, to see their devices and logins, and to manage their activity. Facebook and other identity providers offer similar tools and encourage their users to check their security settings periodically. Clearly, a growing number of users is interested in controlling their privacy, and they're capable of making granular decisions about their online security.

Furthermore, Google, with some other leading technology companies like Microsoft and PayPal, founded the FIDO Alliance, which aims to standardize secure and easy-to-use authentication methods across the Internet.

Meanwhile, another development in account security could leverage the *network effect.* When a hacker gets a hold of a large number of passwords and starts trying them on different sites, those attempts are seen separately and are therefore not identified as an attack. However, if companies were to share security data, using the network effect, these individual logins could be identified as one unified attack and stopped across all participating sites before major damage is done. Gigya introduced this network effect approach in 2016.[34] Facebook also has introduced a network-based security mechanism that aims to connect information from different websites so they can better protect their users.[35]

Could biometrics replace passwords?

Initially, when companies wanted you to *have* something to verify your identity, they would send you a special USB device or a key card of some type. Today, everyone has a device they take with them everywhere, which they consider an extension of themselves. Mobile devices are sophisticated enough to provide essential functionality when it comes to authentication. As mentioned before, the most common form of second-factor authentication is sending a text to someone's phone. Another common practice is using a dedicated authentication app that identifies the device. However, as much as most people act like their phones are a part of them, they actually aren't, of course. Like anything else a person has, the phones can be lost, stolen, or broken. But what is always with you? Your body.

Biometric identification—fingerprints, facial recognition, and voice recognition—are available and in limited use today. Like every other form of authentication, biometric identification has its promises and its challenges. For example, so far biometric use is limited because it requires hardware. Not every device has a camera or fingerprint reader. A bigger barrier though is that, as of this writing, biometrics aren't mature enough for reliable recognition over vast databases. Right now, devices that can be opened by a fingerprint are matching your fingerprint to a database of one: it's either yours or not. Although some flaws remain, whether with fingerprints[36] or facial recognition,[37] there are significant advancements, like Apple's FaceID with iPhone X.

Currently voice biometrics are increasingly being adopted as well,[38] because customer service so frequently is provided over the phone, and every individual has a specific voiceprint that includes his or her voice, accent, and speech patterns. Voice biometrics can serve as a second form of identification for anyone calling from a known phone number. When people call from their mobile phone or type in an account number, a database can identify who they are and then verify their voiceprint from the database of one person who is associated with that phone number or that account number. Of course, this form of identification won't work if a person is using a phone that isn't associated with her voiceprint. Additionally, voice biometrics face challenges when people's voices change. In situations where

people call only a few times per year, changes in their voices over time become problematic for today's technology.[39]

As the technology of authentication advances, so does the technology of hacking. Available technologies could enable today's hackers to create voice and facial matching to outsmart biometric authentication software.

Yet another significant and ongoing obstacle to using biometrics is people's fear of having their privacy breached. For example, a school lunch program wanted to use fingerprints to identify students to speed up the process of distributing lunches, but local citizens decided that fingerprinting was an invasion of privacy, inappropriate for the minor benefit of having a faster lunch line.[40] Furthermore, depending on the types of biometric systems, certain individuals may wish to decline to participate based on religious or cultural aversions.

As governments increasingly implement biometric identification for regular citizens, both security and privacy concerns arise. For one thing, in some cases, individuals can't opt out of biometric identification. When you're standing in the customs line at an airport, you aren't in a position to refuse to be fingerprinted. Or in your home country, you may be required to use biometrics to get government benefits or even to file taxes. Unlike companies that offer customers different security options, when governments decide to collect data, people have no choice; they must comply.

After a government holds biometric information, it can't be retracted. A person can be identified for the rest of his or her life based on that biometric data. No one knows how a government may link information between authorities. In some countries, this cross-referencing occurs between agencies, providing extremely detailed profiles of people across different services. Although you may trust your government, not everybody in every country can. As history has shown, governments undergo changes, with one administration or regime proving less trustworthy than another. But when your government stores biometric data, that data can't be changed. Unlike a name or a cell phone number, your biometric data is permanently with you and permanently on file.

When biometrics data are stored locally, such as a fingerprint or face recognition login on your cell phone, biometrics can't stand alone as your only means of identification. If you lose your devices or other security breaches occur, you still need your passwords and other information to authenticate yourself. In the short to medium term, you'll still need passwords, physical forms of authentication, and security questions to authenticate yourself.

The Future of Authentication Is Like the Past of Authentication

As I started this chapter, I discussed the seamless authentication of the real world: you see someone and you know who he or she is. This real life authentication is an ongoing process. At every moment, a person continues authenticating his or her identity with his or her speech, body movements, memories of you, and interactions with you. If anything doesn't match up, you become suspicious. If that person's voice doesn't match, you would probably feel puzzled. If that person starts espousing spiritual revelations or opinions that don't match past views, you might ask what has changed in his or her life. In other words, although I'm speaking about authentication as a one-time activity, it isn't. Authentication is a moment-by-moment process, and that's what you can expect to have happen in the area of digital authentication as well.

Already authentication combines behaviors with passwords and biometrics. Data is recorded about what times of day you log in, how frequently you use your devices, how fast you swipe and type, how loudly you speak, what devices you own, where you're located, and what types of words you use. Even if you gave someone a password you use every day, she probably wouldn't type it in as fast as you do when you sit down at your computer. If she were holding your phone, she wouldn't swipe the same way or tap with the same intensity.

Just as you know that a man is likely to have a scruffy chin at the end of the day, even though he was clean shaven in the morning, as

visual identification and cameras improve, a biometrics device can know that as well. Just as you know that your friend can't be at home when you just dropped him off at an airport, so can identification systems. In these ways authentication will morph into an ongoing, moment-by-moment process. If someone suddenly performs an action from a location where her mobile device isn't located, that's immediately suspicious. Alerts can be triggered by any action that seems incongruent with all of the biometrics, behavioral, and historical data about that person.

A current initiative, Google Trust API, is attempting to use multiple signals—not just the possession of the phone or a text from the phone, but the behavioral signals a phone can detect—for authentication. With this authentication method, an Android device would issue a trust score for you, based not on a password but on your personal usage traits.[41,42]

Ultimately, your digital authentication will be transparent. At every moment, the devices you're using will know who you are. At every moment, multiple signals will be correlated, just like in real life. Companies such as SecuredTouch are developing technologies that track *behavioral biometrics,* patterns of behavior that identify you as an individual.[43] Although some people might think of this as creepy, these authentication methods simply replicate what we, as humans, already do. We notice that a friend stopped eating meat or drinking coffee, we know their reactions to our opinions, we know how they act (and smell) before and after a workout, we know they text with their thumbs or fingers, we know how they walk or run, and we know how they dress for specific occasions. As humans, we remember all of this data, as well as things computers don't yet know how to detect, and we use it to identify one another. We also communicate to one another about these changes: "Did you notice something different about so-and-so?"

Usually, nobody is impersonating your friend in real life, but your digital identity is more likely to be impersonated. Creating systems of authentication that leverage more of this behavior data could one day make your life significantly easier and safer, saving you time and making it much harder for anyone else to use your identity.

Say that everyone is enjoying greater security thanks to new systems of authentication. But what happens if you can't authenticate? You may forget your password, you can lose a device, and you can even change physically, so voice recognition, for example, could be affected by something as simple as cheering too loudly for your soccer team or being ill with a bad cold and losing your voice for a few hours. What if someone has an incident that radically changes his behavior? What if a stroke or a car accident changes his body? If an authentication system has difficulty recognizing someone based on his stored identity data, a backup would be needed. For this reason, passwords may never disappear completely. Under certain circumstances, a password, written down somewhere, may be the only way to retrieve your identity. But as I've stated, passwords are problematic. So can a truly sustainable authentication system be created?

In early 2017 Facebook released its delegated account recovery, an open-source solution that can be used by any company, such as Twitter or Google, to confirm a user's identity if a password or other authentication method fails. This solution recognizes that email and text validation are fundamentally flawed ways of validating an identity.

Delegated account recovery comes close to a social effect because it uses the authentication resources of other parties with whom you have a relationship. Although it's an interesting step, it doesn't yet resolve any of the underlying issues with the three pillars of authentication: something you have, something you know, and something you are.

So if you might forget things, lose things, and change your physiology, what is a sustainable method of proving who you are? Sampling DNA isn't a particularly easy methodology, and it still carries with it the drawbacks of biometrics and privacy concerns. The method that seems to have the most intuitive potential is the old-fashioned method: having people who know you vouch for you.

Through a trusted network of authenticated individuals, you can ask for validation when you forget a password or identifier. By creating a friend-based authentication recovery method, you can create something reliable, secure, and sustainable through any kind of change in your possessions, knowledge, behavior, or body. Facebook, being the

ultimate social platform, has shown some attempts to provide such functionality.

Friend-based authentication can be a reliable yet nonbinding alternative to biometrics. It could become the solution for the authentication recovery problems everyone will face more and more as identities get completely digitized. Through a network of people affirming each other's identities, like the small village whose inhabitants vouch for each other, everyone's acquaintances in the global village could be the most important protectors of our identities.

Endnotes

1. https://www.ynetnews.com/articles/0,7340,L-4959456,00.html
2. Assaf Mischari (security expert, head of research at Team8, Former CTO in the cyber division of the IDF's Technology and Intelligence Unit 8200) in discussion with the author, May 2017.
3. https://www.wired.com/story/yahoo-breach-three-billion-accounts/
4. https://www.wsj.com/articles/u-s-suspects-hackers-in-china-behind-government-data-breach-sources-say-1433451888
5. http://edition.cnn.com/2015/06/22/politics/opm-hack-18-milliion/index.html
6. https://www.consumer.ftc.gov/blog/2017/09/equifax-data-breach-what-do
7. http://timesofindia.indiatimes.com/india/25-of-Indian-births-not-registered/articleshow/12104158.cms
8. http://trivia.serendip.in/trivia/lal-bihari-mritak-and-association-dead-people
9. http://www.dailymail.co.uk/news/article-2427313/Benefit-fraud-Emilio-Brunetti-paid-alcoholic-neighbour-pose-dying-father-16k-con.html
10. https://www.baytoday.ca/local-news/man-uses-dead-relatives-to-collect-government-benefits-356412
11. http://www.seattletimes.com/seattle-news/special-reports/she-stole-anothers-identity-and-took-her-secret-to-the-grave-who-was-she/
12. http://www.thepresstribune.com/article/9/28/16/roseville-woman-serve-145-years-mortgage-fraud-identity-theft
13. http://www.timesunion.com/local/article/Florida-man-gets-6-months-for-stealing-dead-10686091.php
14. http://www.nydailynews.com/new-york/queens/queens-sanitation-worker-posed-dead-twin-welfare-benefits-article-1.2954821

15. https://www.cylab.cmu.edu/files/pdfs/reports/2011/child-identity-theft.pdf
16. https://www.youtube.com/watch?v=bjYhmX_OUQQ
17. https://www.dhs.gov/xlibrary/assets/usvisit/usvisit_edu_10-fingerprint_consumer_friendly_content_1400_words.pdf
18. http://www.straitstimes.com/singapore/courts-crime/hit-and-run-cases-slowing-down-thanks-to-videos
19. https://www.yahoo.com/news/hackers-hold-entire-hotel-ransom-213915523.html
20. https://www.wired.com/2016/08/jeep-hackers-return-high-speed-steering-acceleration-hacks/
21. Hanan Levin (security expert) in a discussion with the author, July 2017.
22. http://www.linuxjournal.com/content/passwordping-ltds-exposed-password-and-credentials-api-service
23. http://www.csoonline.com/article/3152787/data-breach/black-market-medical-record-prices-drop-to-under-10-criminals-switch-to-ransomware.html
24. https://www.symantec.com/connect/blogs/underground-black-market-thriving-trade-stolen-data-malware-and-attack-services
25. https://gdpr-info.eu/art-20-gdpr/
26. http://www.dailymail.co.uk/sciencetech/article-4125128/The-common-passwords-used-2016.html
27. https://13639-presscdn-0-80-pagely.netdna-ssl.com/wp-content/uploads/2017/12/Top-100-Worst-Passwords-of-2017a.pdf
28. http://www.tomsguide.com/us/password-manager-pros-cons,news-19018.html
29. https://venturebeat.com/2017/04/18/new-password-guidelines-say-everything-we-thought-about-passwords-is-wrong/
30. https://www.wired.com/2016/06/hey-stop-using-texts-two-factor-authentication/
31. http://gizmodo.com/how-hackers-reportedly-side-stepped-gmails-two-factor-a-1653631338
32. https://www.techworm.net/2016/06/hackers-bypass-googles-two-factor-authentication.html
33. https://www.usatoday.com/story/tech/talkingtech/2017/04/29/careful-before-posting-your-10-concerts/101078214/
34. https://www.gigya.com/blog/strength-in-numbers-gigya-introduces-network-protected-identity/
35. https://developers.facebook.com/products/threat-exchange/
36. http://ieeexplore.ieee.org/document/7893784/?reload=true
37. https://www.forbes.com/sites/amitchowdhry/2017/03/31/samsung-acknowledges-galaxy-s8-facial-recognition-security-limitations/#6c326dc21aff

38. http://www.computerweekly.com/news/450298245/Singapore-banks-adopt-voice-biometrics-for-user-authentication
39. https://www.americanbanker.com/news/voice-recognitions-surprise-shortcoming-aging-customers
40. https://www.nap.edu/read/12720/chapter/6#89
41. https://www.theverge.com/2016/5/23/11749938/google-android-password-trust-score-api-io
42. http://www.networkworld.com/article/3074664/security/google-s-trust-api-bye-bye-passwords-hello-biometrics.html
43. https://securedtouch.com/

4

The Value of Relationships

Even a hundred years ago, businesses knew their success depended upon the quality of their relationships. When a customer walked into a shop, the shopkeeper learned her name. Eventually he came to understand her spending habits. Over time she developed a credit history with him. Perhaps the shopkeeper even had a notebook in which he kept customer tabs, recording purchases for payment at the end of the week, the month, or the season. When a customer's child came into the shop to pick up an item, the shopkeeper recognized the child and charged the right customer's account. Thus, shopkeepers developed what today would be called *personal profiles* for their customers. If a customer had bought a part to fix something in his home, the next time that customer entered the store, the shopkeeper would inquire about the repair. Every interaction furthered the connection between shopkeeper and customer. If the shopkeeper behaved within social norms, these interactions resulted in rich, trusted, and loyal customer relationships.

But what about today? Today's customers don't even go into a shop. They're more likely to pull out a device and tap a screen a few times to make a purchase. Where there used to be a shopkeeper getting to know his customers, now devices collect data about their customers. Unlike the shopkeeper, though, the device doesn't know that the customer actually bought that vase as a gift for her mother-in-law's birthday. So instead of inquiring about the birthday party, the

next time this customer enters the web-based showroom, the virtual shopkeeper inundates her with more vase recommendations.

For companies in the digital age, the intent of data collection is the same as the intent of the shopkeepers' personal inquiries a hundred years ago: businesses want to get to know their customers so they can offer more of what customers want and create better user experiences. Though the means of data collection are different today, customer service challenges remain the same. Salespeople of the past had to respect the lines between inquisitive and invasive, between persistent and pestering. So do businesses selling online.

As online marketing has evolved, the rules—not just government regulations, but the ideas of what is socially acceptable—have changed. Just a decade ago, you heard companies talking about "owning the customer," but now industries speak about "building relationships." Businesses haven't become benign, of course, but they know they must address consumers' needs, preferences, and concerns in order to survive.

The Wild Frontier, Then and Now

Today, people constantly engage with businesses online and are generating increasingly more data about themselves. They browse and search more, consume more media, and fill in more forms. As this data is generated, businesses make use of it to create value both for themselves and for their consumers. Companies use data to personalize customer experiences, a marketing strategy that pays. "I've seen a whole range of metrics jump between two times and four times when real time individual personalization is applied," says Charles Nicholls, head of SAP Hybris as a Service.[1]

The Internet has given businesses a power never before available at this scale: the ability to develop deep relationships with millions of customers. Today, businesses can gain intimate knowledge of people's habits and classify users into categories based on their behaviors and preferences. With the technologies available today, as companies collect

more data, each customer can become his or her own segment, a person-specific profile that enables individualized experiences.

Initially the Internet was a wild frontier, with no standards governing marketing practices. Third-party data collection ruled the day. Data brokers collected consumer data from various sources and sold it to businesses for their use. The first attempts to understand customers and to serve targeted advertising were (and sometimes still are) somewhat clumsy, from showing customers advertisements for products they've already bought to addressing emails to the wrong name. As regulations like the General Data Protection Regulation (GDPR), the European privacy regulation, come into play, the wild frontier is being tamed, somewhat. Third-party data collection continues, but it's meeting serious limitations, regulatory and otherwise.

Consumer profiles created by data brokers are often inaccurate and outdated, not to mention that data collected by third-party tracking cookies is device specific, so it can't produce a full picture of users' activities across devices. Further, some browsers now default to blocking third-party cookies. At the time of this writing, Google's Chrome browser doesn't yet default to blocking, but eventually customer pressure likely will push Google in that direction.

As third-party data falls out of favor, companies realize the value of *first-party data,* the personal information they collect directly from consumers. First-party data systems can collect data across devices, through a variety of site or app interactions from the click stream, registration fields, subscription data, social network logins, and more. Using permission-based methods, companies can obtain data points such as interests, purchase behavior, favorite brands, gender, and age.

Despite changes in regulations and customer expectations, the Internet continues to offer opportunities for vast market growth. Although some business people bemoan the loss of the lawless and wild frontier, I see new opportunities arising. Instead of focusing on collecting data without a customer's knowledge, what happens when businesses focus on initiating and growing unique relationships with each of their customers, in which customers offer their information because they understand the value—to them—of sharing?

Get Creative: Using Customer Data

Let me discuss how a few companies are leveraging consumer data to enhance customer experiences and create valued relationships.

Disney's first theme park established the company's leadership in the amusement park industry. These days Disney is leveraging customer data to enhance guests' experiences. With the use of an app and a wristband that tracks data, guests can order lunch and have it delivered to them anywhere in the park. Rides are personalized with screens that welcome guests as they arrive and say good-bye as they leave, and hotel room doors open upon guests' arrival. Guests also experience personalized interactions with Disney characters and gain easy access to photographs and videos taken of them on each ride. As the wristband gets to know guests better, the consumer experience becomes even more seamless. Recently, Disney announced plans for a Star Wars–theme hotel, where data-driven technology will allow each guest to experience a personal storyline that will unfold throughout their stay, aided by in-character Disney cast members.[2]

Clearly, Disney has tapped the power of consumer data, leveraging the information consumers' consent to share. But what are other companies doing?

In the fashion space, with a hyperfocus on targeted marketing, ASOS.com, a global online shop for fashion-loving 20-somethings, is seeing 34 percent year-on-year growth, with a 1.9 GBP annual turnover in 2016–17. The company has more than 15.4 million active customers across 240 countries, offers 85,000 products, and is adding 5,000 new products every week. How does the company compete with the major fashion retailers? By leveraging a much higher granularity of personalization than its competitors. To the customer, the website looks like a hybrid of a fashion magazine and an online retail shop. By tracking both the styles customers browse and the content they read, ASOS comes to understand its customers in the same way a personal shopper gets to know her individual client. Through this approach, ASOS offers a boutique shopping experience to a vast customer base.

In other cases, data collection can help companies zero in on their most loyal customers. For instance, Robert Dawson Scott, head of engagement for Scottish Television (STV), points out that the network has millions of viewers, but approximately 100,000 viewers spend the most time on STV channels. "We think very hard about how we treat those hundred thousand people. We think about how we can make them feel special, because they are special. To us." He adds, "Sometimes, one magic moment in the user experience is all it takes."[3]

Even when companies are at full capacity, they can use personalized data to improve their returns. Craig Ambler, Director of IT for Center Parcs, a network of resorts and vacation homes based in the United Kingdom, says the company's properties are at 97 percent capacity year-round, and the data collected allows the company to maximize each customer's spending. "Some customers come on vacation and eat out for every meal, while others come in with their own food and spend a lot less," he says.[4] Through this data about customers, Center Parcs is able to understand the specific experience each customer wants and to plan his or her offerings, which may include better upselling opportunities.

As people spend more time connected to the Internet, businesses are beginning to understand and explore opportunities for building direct relationships with their customers. These opportunities lie ahead for both companies that sell to customers (B2C) and companies that sell to other businesses (B2B). Finally automotive manufacturers can communicate directly with drivers who buy their cars, through websites, apps, and even the cars. CPG companies offer apps and connected devices to engage with their customers before and after the product has been purchased, and in many cases through the product itself. Even pharma companies understand the potential. I'm now working with a pharma company that is designing a smart device for diabetes patients. The device will collect data about insulin levels and send it to the cloud to be picked up by an app that provides smart recommendations personalized to each patient.

I also work with a luxury watch manufacturer that engages its audience (only customers who purchase the product, in this case) with fascinating content related to the technologies behind the products

offered. EUrail, the European rail company, made a major leap, offering its community an experience, from suggested travel resolutions to social adventure sharing. Today people live in an era when companies offer much more than a mere product or service. They engage their customers in a comprehensive experience incorporating various elements orchestrated and tailored for each individual, spanning over time. In this way customers begin to identify themselves with the brands they engage, which fosters customer loyalty. It's an ongoing cycle that involves value, trust, and exchange.

Value → Trust → Exchange

The collection of first-party data depends on a cycle of value → trust → exchange.

For a relationship with a customer to progress and a customer profile to grow, this cycle must repeat itself, in small iterations, over time.

To illustrate how this works, consider a customer researching accommodations for an upcoming vacation. Josh lands on a travel website and begins clicking his way around, looking up room availability, prices, spa, and on-premise dining options. After several minutes of exploration, the website is able to analyze, at a basic level, Josh's preferences and interests to present him with a relevant promotion: "Sign up for our newsletter and get a spa package with your next visit!" Josh, who is a spa fan, signs up.

This interaction triggers an email offering an upgrade package featuring a free spa treatment, if Josh books a room within a week. Josh was already planning to travel soon, so he books a poolside room. After completing a quick check-out process that requires him to fill in only necessary details, Josh is offered points for the nights he booked through the site's loyalty membership plan. Becoming a member is free and requires filling in just a few more details. The site already has Josh's email address, as well as some other data from the booking process. Josh answers a few questions about his travel habits and chooses a password. He has now signed up for the site's loyalty program. The site now knows that Josh travels about four times a year to Europe

for work and twice for leisure. All this information is added to his profile. After the form is submitted, Josh receives a link to install the site's mobile app. The offer states that the app will notify Josh about relevant attractions during his planned travel.

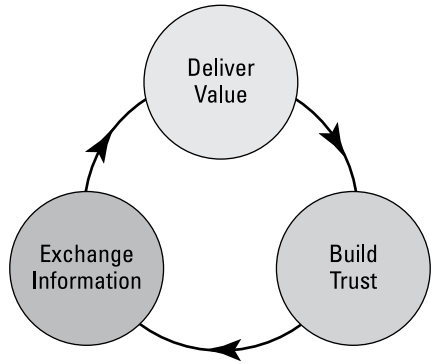


Figure 4-1: *The continuous cycle of value → trust → exchange.*

Now that he has joined the loyalty program, installed the mobile app, and provided key personal details, the site can cater more offers to him and message them through the right channel, be it an email or an app notification. Each time Josh receives an email from the loyalty program, he's asked a question with a simple drop-down answer menu. Over time, the site learns that his preferred activities include "staying in and reading books by the fireplace" and "hiking and biking," and he identifies as a "shopaholic" and a "country club regular." The site can now include conditional fields based on his interests to be filled on-site or through the ongoing email campaign. This might include questions about his favorite retail clothing brands or booksellers in the area.

In one of his following visits to the site Josh is asked if he'd like to connect his social network identity with his site account, so he can later log in to his account from any device using his social network credentials. Josh agrees to connect his Facebook account, and he also accepts the social login prompt that asks to transfer to the site the list of his social likes and interests. From now on, the site will get updates

from Josh's social network profile whenever Josh adds likes or makes profile changes.

Josh's trip has arrived. During his trip, he is prompted with an app notification about a great event that takes place in the city he visits and that fits his interests. Josh confirms that he might choose to attend. One day after the event takes place, the app prompts Josh with a request to write a review about the app in the app store. Because Josh enjoyed the event, he gladly adds a recommendation for the app.

And so the cycle continues: businesses provide value and thereby gain consumers' trust, and consumers become more willing to share their personal information. This rich data enables companies to present increasingly more personalized interactions and offers and to do so more accurately than they ever could with third-party data.

Even though solutions for personalized marketing are far from perfect, the objective is clear: offer the right products and services to the right people, at the right time, and in the right way.

Progressive identity

Many of the systems companies use to manage customer data treat the data as a binary: they either consider the customer as identified or they don't. But in real life, relationships don't work that way. As I previously discussed about shopkeepers of old, and in Josh's case, businesses can come to know their customers' identities progressively, over time, through the customer's journey and the iterations of the trust cycle.

When someone browses a site, the company doesn't know who she is. Cookies may track this behavior, or the company's website might know that this person is clicking on certain pages in a certain order. At this point, the user's identity is made up of an activity pattern and is limited to a specific browser. At some future point the user may write a review, leave a comment, or maybe join an email list. This would represent the first time the person has given identifying information. As this person registers for a service or purchases an item, the company gathers more information, such as an address and credit card number. Of course, the purchase is critical to the company, so companies are

careful not to ask for too much information when someone initially joins a service or makes a first purchase. Over time, however, the site can ask a user for additional information about herself, a practice known as *progressive profiling*, or can ask questions relevant to particular products or purchases, such as "Is this purchase for yourself or for a family member?" A user may be inclined to answer further questions, such as indicating it's for her son and providing her son's age, if doing so improves future search results. As long as the questions are appropriate for the level of trust the site has established with the customer and the customer sees benefit in answering the questions, the customer will be open to sharing additional information over time.

The identity of the customer continues to mature through other systems the customer engages. Customers connect with companies through support calls, chats, and marketing channels—like when they click on on-site ads or on links in emails. They also connect offline, at smart stores and regular stores' points of sale (POS). All this data assists in constructing a comprehensive profile that feeds back to these channels a smarter engagement. Not only will content and marketing become more personalized and relevant, but even support representatives will be able to know, for example, that the customer they're now talking with is an influencer with more than 10,000 Twitter followers and that she never clicks on ads. Given this information, that support call might sound different.

Customer identities progress in other ways as well, as people develop and change. For example, people's family situations change, their locations change, and their preferences change. To optimize the value of consumer data and to comply with regulations, customer identity and access management (CIAM) systems need to treat customer identities not as something to categorize by levels of engagement but as individual buckets of progressively collected information, each belonging to a category of one: the individual customer.

Furthermore, businesses can enrich their customers' identities not only by learning what customers do on websites or apps but also how they interact with smart devices. Today's cars, healthcare devices, and even kitchen appliances record data about customers, from when and

how far they have driven to how often they open their refrigerator doors after midnight.

Contextual personalization

To succeed in engaging with today's digitally savvy consumers and to deliver them a personalized and relevant customer experience, brands need to become contextual in their interactions. Customers today choose their own path to purchase a product or service, which makes every journey unique and a challenge that drives the need to move beyond general segments and to understand each individual's preferences, intent, and purchasing triggers. In addition, to stay relevant with messaging, real-time context is key. Many organizations still overemphasize preexisting data, which oftentimes is hopelessly outdated. For example, as a consumer, I'm only so often interested in purchasing a new TV set or laptop computer, still the practice of being followed around by product recommendations and ad campaigns weeks after the actual purchase persists.

Availability of real-time context would extend insights beyond what customers have done in the past and what their future propensities may be, to provide a clear view of what they are doing and wanting right at that moment. This would include on-site session and browsing data, to derive intent and interest signals and to gain insights into where they are in their purchase journey. It would include interaction data from mobile apps and IoT devices for an in-store iBeacon tracking, but also real-time engagement on social channels. However, real-time context doesn't stop at the customer. Other factors also need to be accounted for, including content and product-related context (for example, currently trending products or videos), but more importantly location-based data, like a customer's proximity to a store or the fact that a customer is at a sports event. Even referencing the current weather in a geolocation can drive incremental success of personalized marketing.

Challenges in Customer Identity

The operation behind leveraging the power of customer data is complex, to say the least. It involves multiple systems used within a company, as well as the systems external vendors use. Large organizations that run multiple properties and across multiple geographies experience even further complexity, because businesses must comply with the unique sets of regulations established by customers' home countries. The multitude of data collection methodologies adds yet another complication into the mix. A customer may connect with a brand through a purchase in a real store, a visit to a website or app, or a call to a support center. This customer expects a consistent experience, and regulations expect consistency as well, across these various means of connection. However, according to Accenture, 78 percent of customers receive a fragmented experience as they move channel-by-channel.[5] To make things more complicated, many customers choose to register using their social identity from Facebook or Twitter, and these Internet services have their own rules about what can and can't be done with their customer data.

Envision a company that manages different websites in different countries. They might not recognize a traveling customer who logs in from multiple locations. Such a customer might find herself creating duplicate, segregated accounts, leading to a broken experience. Consider a company that acquires another company or upgrades a website, which often results in overlapping data for a particular individual. Or, what if a company with multiple consumer brands wants to sustain a single customer view? For example, someone who is a fan of Nescafé may also have entered a sweepstakes at the KitKat site, so now Nestlé has multiple accounts for this individual. Combining such accounts isn't so simple, because most consumers engaging with each of these brands individually don't think of themselves as building a relationship with a conglomerate called Nestlé. Yet ideally a system would allow Nestlé to maintain one customer record, whereas the customer interacts with each set of data separately.

Companies have an interest in getting a full picture of consumer' behavior over multiple areas, but this effort comes with significant challenges. For example, a parent may purchase a certain type of baby formula and then progress to baby food while, in parallel, a site that sells toys could sync its toy recommendations accordingly. For a company to achieve this, three challenges come up. First, the system would have to be sophisticated enough to make the connections. Second, companies would have to ask customers explicitly for consent for cross-brand promotion. Third, companies need to be sensitive enough not to cross the line between helpful and intrusive, as in the famous story from 2012, when Target learned that a high school student was pregnant before her father did.[6]

Under the real pressures of consumers' expectations, marketplace competition, and regulatory fines, businesses need to face these multiple and overlapping challenges head-on, proactively, and with ingenuity. But how?

How Smart Companies Handle Identities and Profiles

The myriad challenges associated with customer data management have companies searching for holistic identity management solutions.

Historically, identity was managed by an organization's IT department, as it addressed security needs. Identity and access management (IAM) solutions enable companies to maintain a single identity for employees across a multitude of systems within an organization. Although IAM systems were born as security solutions to address a problem, customer identity and access management (CIAM) systems are designed to leverage an opportunity—creating a long-lasting and trusted relationship with the company's customers. CIAM systems are developed from the business and marketing point of view and aim to help companies address the complexities involved in managing customer identities. Where employee identity systems deal with tens of

thousands of records, CIAM solutions must handle tens of millions of customers with occasional spikes of thousands of logins per second.

Although regulation is an important driver behind CIAM systems, consumer demand is also a major factor pushing companies toward CIAM solutions.

According to Marcus Ruebsam, an industry expert and a senior vice president at SAP Hybris, the trend in 2019 and 2020 will move toward an even more holistic approach that puts the customer profile in the center of an organization. "The combination of identity, consent, and profile can become the ultimate approach for managing customers in the coming years," he said.[7]

A business that takes a comprehensive approach to customer relationship management will need to incorporate the following capabilities:

- Maintain a single record from unknown customer to known customer. This golden record is born at the very first interaction between the customer and the organization (entering a store, visiting a website, installing an app, turning on a smart product, etc.), and it grows as the customer interacts with the business through different channels and using various devices.
- Support data and event processing from high volume channels (Web, IoT, etc.) at an individual level, made possible by modern event-streaming technologies with capabilities for real-time metric calculation.
- Avoid duplications of identities using deterministic or, when needed, probabilistic matching.
- Keep data hygiene by correction and clearance.
- Support all the phases of progressive identity, as mentioned earlier in the chapter, including activity data, light registration, full registration, social login, and progressive profiling.
- Store and index all data in a centralized database.
- Create actionable insights from the data gathered, in order to activate personalization of the different systems with which the customer interacts. Insights should process real-time activity and behavioral data in order to address a contextual experience.

- Offer predictive modeling.
- Support relationships between customers (e.g. parent and son) and between customers and devices (e.g. car and its drivers).
- Support the orchestration needs of enterprises with a global presence, multiple brands, and properties.
- Connect with all organization systems that either generate or utilize customer data.
- Provide all security elements related to authentication and authorization, including single sign-on across properties.
- Provide full data governance and consent enforcement capabilities, including the features needed to comply with worldwide regulations and policies.
- Provide a clear map of where customer data is stored in the organization, how it's used, and the flow of procedures related to privacy regulations (e.g. if a customer wishes to be forgotten, how will the flow of executing such a request occur, including dependencies and actions that will take place in later time?).
- Be simple enough for the business user (e.g. marketer) to operate.

The gold standard of customer data management would be to reach the level of intimacy of the old-time shopkeeper who knew each of his customer's personal habits, histories, and families, and who used this information to offer personalized services and products that improved his customers' lives. Today's technologies enable ever greater possibilities for this kind of personalization. Advanced CIAM and customer data platform (CDP) systems can integrate a customer's activity, site registrations, purchase history, email-based light registration, profile data, social media attributes, communications preferences, and consent under a single record. A unified customer profile will also sync data to and from marketing, CRM, revenue, and other systems the organization is utilizing. Although information may have been collected through different systems and may be used for different purposes, companies now have the tools that can identify a single person and recognize that a diverse range of information

belongs to that one individual. Such tools also can give customers control over their data, which is their lawful right in many countries. Further, CIAM, consent, and some CDP systems today can allow customers to log in to see precisely what data a particular company has collected about them and what permissions they have granted to this company. Customers can even withdraw permission or remove data that they don't want a company to use. Among other things, such systems also ensure that, within the company possessing the identity data, only authorized employees have access to customer data and only at appropriate touch points.

The average company holds customer data spread across twelve different systems, with some enterprises using as many as forty or even more, according to Charles Nicholls, head of SAP Hybris as a Service.[8] Consolidating data under one management system is an essential step for companies that want to provide a higher level of personalization to their customers. Without the right systems to manage this data and flow it effectively, customer experience will remain broken.

Endnotes

1. Charles Nicholls (head of SAP Hybris as a Service) in a discussion with the author, December 2017.
2. `https://techcrunch.com/2017/07/15/disney-is-opening-an-immersive-star-wars-hotel-where-each-guest-gets-a-storyline/`
3. Robert Dawson Scott (head of engagement for Scottish Television) in a discussion with the author, June 2017.
4. Craig Ambler (director of IT for Center Parcs) in a discussion with the author, June 2017.
5. `https://blogs.oracle.com/marketingcloud/modern-marketing-essentials-guide-to-data-management`
6. `http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html`
7. Marcus Ruebsam (senior vice president, Global Head Solution Management, SAP Hybris) in a discussion with the author, January 2018.
8. Charles Nicholls (head of SAP Hybris as a Service) in a discussion with the author, December 2017.

5

Identity Data 101

As a co-founder of Gigya, I've talked to business leaders around the world about collecting, securing, and leveraging consumers' data. Contrary to customers' fears, I've observed that reputable companies tend to be more protective of their customers' data than many consumers think they are. In part, companies' caution is born of their own interests—they understand the financial implications of breaches of trust. It serves their bottom line to serve the best interests of customers. Yet, in interview after interview, I have found that companies' concerns about data management stem not so much from a fear of inadvertently violating regulations, but more from a desire to maintain long-lasting relationships with their customers. Rightly, businesses see trust as the basis for these relationships.

As such, it's a company's imperative not just to comply with regulations like the European General Data Protection Regulation (GDPR), but also to treat identity data the way customers want their data to be treated. Companies who see regulations as a burden will likely lose customers to competitors who use cutting-edge data management practices to reach beyond compliance and enhance their relationships with their customer base. To gain and keep a competitive edge, companies will need to be creative and careful with their customers' data. Starting in this direction, allow me to explain the nuts and bolts of customer identity data as it's collected, used, and managed today.

The Evolution of Data Use

Over the years, companies have learned how to extract increasingly more value out of the data they collect from their customers. This data usage happens on these three levels:

- **Basic:** *Basic data* use enables companies to target marketing and services based on specific data points. For example, if you look for blue running shoes, you'll be categorized as a sports enthusiast and presented with targeted advertisements based on this category. Also, if during site registration you provide your gender and age, doing so puts you in defined categories.
- **Inferred:** *Inferred data* use occurs at a level deeper, using mechanisms to analyze customer data and come to conclusions about customers, thereby creating new data about them. For example, a collection of data might suggest that you belong to a certain religion, socioeconomic class, or political party.
- **Individualized:** *Individualized data* use, the ultimate usage level, allows companies to treat each customer not as a member of a larger category, but as a category of one, an individual. By analyzing a wide range of data points collected from you, over time, and mainly in real time—because context is a critical factor—companies can perfect your offerings and experience.

As individualized data use has become more advanced, inferred data use has become less important. Rather than profiling people into set categories, individualized data use looks at people's behaviors in and of themselves. For example, algorithmically, it's almost as easy to program a computer to associate the purchase of diapers with the purchase of baby toys as it is to infer that people belong to a category called parents of babies. This particular behavior—buying diapers—can trigger an algorithm to show special offers for toys without ever categorizing the customer as a parent. The kinds of toys a person will buy or the changes in sizes of clothing they purchase will follow a

predictable arc, whether or not you teach the computer that there is something called a child who develops in a predictable pattern.

In fact, among the three forms of data use, individualized data use has become the most promising means of forming long-lasting, mutually beneficial relationships with customers.

How Far Should Personalization Go?

Joseph Turow points out in his book *The Aisles Have Eyes* that retail stores have been notorious in profiling customers throughout the history of retailing in the United States, explicitly stocking inferior items in some neighborhoods and treating customers differentially depending on what they're wearing, their gender, or their skin color.[1]

Turow suggests that the data companies are collecting today can be used similarly, and he envisions a future where a customer checking out at a physical shop could notice the person in front of her is purchasing the same item at a different price. What customer wants to expose herself to this kind of discriminatory pricing? Today, differentiated pricing is difficult to get away with in a brick-and-mortar shop. As cashiers and price tags are rendered obsolete[2] and technology enables consumers simply to pick out their items and be charged for them upon leaving the store, differentiated pricing could happen without customers' knowledge.

Now, someone could argue that differential pricing has been, and in many places still is, the de facto business practice. As any business-to-business company knows, many goods and services are traded at differential pricing worldwide. In fact, as Robert Hendrickson writes in The Grand Emporiums, fixed pricing is a relatively new and distinctly western concept, first introduced by retailers in the United States in the 1840s. In markets where fixed pricing has become the norm, a consumer won't abide being charged $10 for that item in her grocery cart when the shopper ahead of her was charged $7. As a matter of fact, to circumvent rampant price discrimination in the airline industry,

many savvy Internet shoppers now search for flights and hotels on incognito browsers, because they suspect online services are spiking prices on them after they start their searches. Here's a tip: don't use your brand-new iPhone or Mac when performing your next travel research. Vendors know the type of device you use and could infer your sensitivity to pricing.

The practice of using data to differentiate prices has been reported for almost twenty years, and it came under scrutiny of the White House in 2015.[3]

Not All Data Is Born Equal

As companies collect and store more data to grow their relationships with customers, they shouldn't overlook these questions: What kinds of data are most *valuable* to the company? and What is the *cost* of obtaining that data?

Companies collect three types of data:

- Data that helps them understand who the customer is, what he likes, and how he shops
- Operational data, such as credit card information and shipping addresses
- Data that enables communication with the customer, including email addresses, phone numbers, and communication preferences

Companies might presume that more data represents more value, but not all data is equally valuable. Different types of data are uniquely useful to different types of companies. Take, for example, someone's full name. This is both one of the most personal pieces of data to a customer and one of the least useful to a company. From a customer's point of view, providing a full name connects all the data he's shared to his real self. Therefore, a full name would be harder to collect than, say, an arbitrary user name. So, a company must consider, is it worth the risk to the company—potentially losing a customer who doesn't

want to share her full name, knowing it will be associated with her behavioral data—to ask for this? From a company's point of view, a customer's last name has little value, unless the company needs to ship something to the customer. Full names aren't indicators of where a customer lives, what he buys, or who his friends are. A full name isn't even a unique identifier. So why risk the expense of gathering a full name if the return isn't valuable?

In addition to considering the value of data, companies have to consider when to ask for each piece of information. The most sensitive point to collect data is at the initial registration point, as the customer establishes a two-way relationship with a company. At this point, the customer is at a learning stage in her relationship with the company. Maybe she's downloading a free e-book or accessing an exclusive offer. Asking for too much data at this point can cause her to abandon the site, so a company needs to consider most carefully exactly what it needs in order to provide its service. Ideally, at this early stage the company will collect only the bare minimum of information necessary to ensure the quality of its service and to establish ways to identify and contact the customer.

The values of different types of data vary from company to company, and the cost of data—what it takes to get users to share their data—fluctuates as well. When requesting data, companies need to consider both the value and the cost.

In my experience, many companies collect much more data than they need. It's not just legislation that's driving companies to think harder about what they collect and how they use data; it's their commitment to service and to customers. Industry expert Ian Glazer says, "You don't have to overcollect to have an omnichannel experience. Sometimes, I can delight a customer and know nothing about him other than his phone number or his Instagram login."[4]

Identity Types

Companies can build trust with consumers by offering data management controls, transparency, and well-timed requests for information.

But how do businesses come to trust the identities of the people using their apps and sites?

Currently companies use a variety of methods for confirming user identities, each offering access to different types of identity data and each characterized by different levels of trust. These identity types include the following.

Unverified

Unverified identities enable a user to enter a site or use an app without any proof that he is who he says he is. For example, a person can register with a made up name and email address. Over time, as behavioral data attached to this name and email address accumulates, a company can begin to "know" this user, even if he never provides his real name.

Verified

Verified identities are authenticated at a variety of levels, such as through an email address, credit card information, phone number confirmation, or real-life information, such as a physical address or a verifiable human who actually shows up to use a service.

Federated

Federated identities are identities that originated somewhere else but can be "imported" or used by allowed entities. Social login is an example of identity federation, a more recent means of obtaining a verified identity, by which customers consent to allow an already trusted identity—like a Facebook account—to be "borrowed" by another business for verification and sometimes for data use.

When Facebook first introduced this feature, it became an immediate hit, quickly adopted by many websites. Why? Well, social login benefits all parties in these ways:

- Users benefit from the convenience of social login. Website visitors feel inconvenienced when they have to create new accounts, choosing new user names and passwords for every company they engage.

- The borrowing companies benefit from social login because it offers a frictionless way to gain access to a rich and trusted customer identity.
- The company providing social login, such as Facebook, benefits because customers who use social login become more committed to their Facebook identities. In addition, Facebook gains information about the sites a customer is using, which informs Facebook's advertising practices. In fact, the value of social login is so significant that soon after the Facebook launch, almost every other major player followed suit, including Microsoft, Yahoo, Twitter, LinkedIn, and overseas companies such as Wechat and VKontakte.

Despite the popularity of social login, some consumers are reluctant to use their Facebook login because they worry that borrowing companies will gain access to too much personal data or that a borrowing company might attempt to share to their feed without permission. Therefore, borrowing companies have begun to offer more than one social login option. This gives users an element of control, because they may choose to use their Google or LinkedIn persona to log in with some websites, while with other sites they may choose to use their Facebook or Twitter login instead.

Federated identity may bring an additional advantage. According to Assaf Mischari, former Cyber CTO at the Israeli cyber intelligence unit 8200, "With all the data they have about you, Google is probably a much more secure login than your bank login system."[5]

Identity federation practices are expanding beyond social login. For example, BankID is an identity used by all Norwegian banks and public digital services and an increasing number of enterprises in a range of different sectors. The first Norwegian customers were issued a BankID in 2004. At that time, the Norwegian banking sector had been working for four years on developing a joint infrastructure. Today, 3.7 million Norwegians have a BankID, and more than one million have BankID on mobile.

A consortium of German banks has also announced a cooperative effort that will enable the federation of bank-verified identities.

At some point, even governments may begin to offer a federation of digital government ID.

Reputation-based identities

Reputation-based identities data plays another important role for online marketplaces. Beyond the initial identity verification, marketplaces like eBay, Amazon, and Etsy collect reputation data about their sellers, establishing a degree of trust with shoppers. Sites like Airbnb, Uber, Lyft, and Gett take reputation data a step further, allowing both service providers and consumers to review each other's reputations, providing a level of confidence and safety on both sides of the market exchange. Naturally, people consider their reputation an important part of their identity, so these ratings encourage positive behavior. Marketplace sites and apps depend on this reputation data to protect both buyers and sellers and to influence their behaviors.

Even when a company's relationship with a user begins with only an email address, accumulated reputation data can develop into a trusted identity. When a user simply offers an email address and gets a newsletter, the company holds minimal data about that user. But if the user logs in frequently, reviews products, and earns reputation points from others, even if the name associated with the account isn't real, the authenticity of the data improves. The company may still not know the user's real-life identity, but it has a high level of assurance that its actions as a company will positively influence the owner of that account.

The shared accounts challenge remains

Although identity types continue to develop, certain challenges remain, particularly concerning shared accounts. A family making purchases on Amazon typically uses one account. With multiple shipping addresses, a family could even live in different locations (such as when a child goes off to college) and share an account.

On sites like Upwork, the employer and employee may never meet one another. "I have this amazing artist in India," said one business

owner, "but sometimes I worry that it's not really a woman I'm working with, but some guy who has a bunch of women who work for him at a lower rate than I'm really paying." In both the Amazon and Upwork cases, you have reputation and identity data associated with an account, but the account might have multiple users. In other words, between people's multiple identities and shared identities, even a "trusted" or "verified" person still isn't precisely identified, making it difficult for companies to target accurately their offers, products, and services.

Keeping It Appropriate

Anyone who has raised a child understands the social learning curve. Most parents have experienced that embarrassing moment when their kid asks the wrong person the wrong question at the wrong time (often in public and too loudly). Businesses also navigate social expectations when gathering information and using customers' identity data.

All businesses want to develop beneficial relationships with their customers, but not all consumers want relationships with the businesses they patronize. For example, consumers expect different levels of interaction and sharing with companies that deal with health and money than they do with companies that deal with entertainment. Customers are willing to record a tremendous amount of personal data and share it in environments that offer personal assistance, like a budget management app. Here customers share highly personal data about every purchase they make, and in response they expect the app to offer highly personalized recommendations about their spending habits. Some brands access a deep level of customer interaction because consumers feel emotionally attached to the brand. For example, customers see diehard brands like Harley Davidson as a part of their identities. However, customers develop almost no relationship with a brand that sells, say, lawn chairs, and they don't see a need to expose any information about themselves to make an informed choice about lawn furniture.

Therefore, companies need to understand what kind of relationship consumers expect to have with them, and then to act in alignment with the consumers' expectation of that relationship, or to build a strategy that will invite a more intimate relationship with customers. Consider a customer who visits an online furniture shop looking for a sofa. He might not choose to share much personal information with the company initially. But if the site offers valuable design advice while he's there, perhaps he would sign up for a newsletter to continue that connection. Consumer relationships are similar to friendships you develop through a hobby. Say you met a friend through cycling. You see that friend at cycling events or on the trails, and you talk about cycling equipment. You don't need to share anything else with that person, and you don't expect that person to share anything else with you. You're happy he told you all about his power gel, and you might seek his advice about, say, bike maintenance issues. Sometimes, these relationships develop into deeper friendships, but most of the people you meet in hobby venues, due to the limiting and narrow context, remain friends only within a limited context.

As long as businesses keep within the bounds of the topics customers share with them and they provide relevant content and advertising, most consumers feel that a particular business is like that hobby-friend: generally benign and a good source of information about the hobby. But just as in real life, if a company violates trust in some way, like asking for overly personal information, sending irrelevant marketing offers, or failing to provide value in exchange for information, customers may never move to the next step.

The Creepy Factor

Whether they're doing a good job of targeting or not, companies are collecting data online, and consumers are accustomed to it. Most people recognize their browser activity is tracked and that the information they share with companies is stored and used to suggest additional services and products. As long as the online experience is helpful and not creepy, consumers generally appreciate the improved

experience this data makes possible. In that context, it's important for businesses to understand how to avoid behavior that makes customers uncomfortable.

Although most people know what creepy feels like, Fatemeh Khatibloo of Forrester Research offers a definition and helpful guidelines: "Creepy behavior violates social norms or does something where social norms don't yet exist," she says. "Where no social norms exist, even benign behavior feels creepy. The first time a voice-activated system told you 'Please wait while I connect you,' it probably felt a bit creepy, because you know there is no 'I' on the other side of the phone."[6]

Khatibloo categorizes creepiness into three types of violations, each with different implications:

- Perceived misuse of personal data
- Data hygiene errors
- Invasive or aggressive tactics with no off button

Even though perceived misuse of personal data ranks most damaging among these violations—people just don't trust a company that seems to know something it shouldn't about them—any one of these offenses can erode a customer's trust in or loyalty to a company.

A few examples of these violations have reached the news headlines. One embarrassing (on the company end) and painful (on the customer's end) snafu resulted when an advertisement was mailed to a bereaved father, addressed to "father whose daughter was killed in a car accident," instead of to his name.[7] Perhaps that information was entered into a computer by a sympathetic customer service representative who wanted to make sure the father was treated with respect given his loss, but it turned out to be an offensive fiasco when the automated system typed it up in a letter.

For companies using marketing automation tools, preventing such errors can be difficult, because computer algorithms don't have social awareness. Algorithms follow a customer's behavior and make preprogrammed suggestions based on that person's activities and similar behaviors of other people. The tracking algorithms may notice that

when someone buys a computer, she also buys a bag or sleeve, and it may notice that at certain times of the year, people in San Francisco buy bags with the 49ers logo. So it may send a customer who fits this behavior pattern an advertisement for a red and gold computer sleeve. The algorithm also might see that customers who purchase computers with a specific type of graphic cards also are fans of a specific brand of computer game. Customers don't know why they got a black "Eat, Sleep, League" T-shirt advertisement when they purchased a computer, and they feel a bit creeped out that the system knew they wanted one.

Pattern recognition can be incredibly accurate, but it has its downfalls, too. These systems don't distinguish the social implications connected to a Neti pot versus a cooking pot, unless they're instructed to do so. Computers don't know it's impolite to tell people to stick something up their noses. A human makes that judgment call easily, but a computer still doesn't know how to figure out. Hence awkward situations arise.

Companies need to be aware of the creepy factor when they track customer behavior in their physical stores as well. According to a recent study, 74 percent of retailers are using technologies in-store to track customers with 27 percent using facial recognition technology.[8] Many are using in-store Wi-Fi and geotracking to map customers' journeys through their brick-and-mortar shops. Others, like supermarkets, use scan-it-yourself technology to understand user behavior. In these markets, customers can carry wands to scan their groceries as they place them in their bags or shopping carts. At checkout, the shopper simply replaces the wand, and all of the items are calculated for payment. This technology benefits the company by providing precise information about the consumer's trip through the store. Customers, on the other hand, enjoy the seamless shopping experience.

Generally speaking, customers have become comfortable with technology that they perceive as improving their experience, like the wand or an online grocery outlet that remembers their regular shopping list. Many consumers appreciate technology that helps them decide what to buy or recommends products in store. However, customer awareness about data collection varies widely. Many custom-

ers realize their loyalty cards collect their data, yet fewer understand how companies are using facial-recognition technology.[9] Even though in-store facial recognition doesn't reveal the identity of a person (yet), many businesses are using it regularly to identify how many people come into the shop, whether they arrive alone or in a group, and other such details.

How would customers feel if they knew facial-recognition technology was being used to profile them? As *The Sun* reported, a pizza place in Oslo, Norway, quickly learned that customers don't like it.[10] Unbeknownst to random passersby, an electronic billboard outside the restaurant had a camera built into it. The camera recorded people walking by and changed the onscreen advertisement based on their profiles. This secret was revealed when a computer glitch caused the profiling text to appear on the screen. A patron walking by read "male–young adult," "attention time," "smile," and "glasses." According to the article, the billboard showed photographs of the restaurant's meatier dishes to men, whereas women were shown images of salads. The patron who noticed the glitch snapped a picture of the billboard and posted it online. The incident had triggered the creepy factor, and news of it went viral.

On one hand, yes, the stealthy nature of this endeavor does seem creepy. On the other hand, this profiling is akin to the kind of service people have always gotten. When you walk up to a restaurant counter, the person can see your gender, age, and facial expression. If you frequent a restaurant regularly, the waiter may know your preferred food and drinks, and then offer that to you before you order. In fact, many people expect that kind of service.

As businesses automate more and more of their services, automated recognition will be a necessity, simply to continue the level of service customers expect. New technologies will enhance these abilities. For instance, in 2017 Facebook was granted two patents that assist in analyzing a user's emotional state. One of them, Techniques for Emotion Detection and Content Delivery, utilizes a device camera to analyze a user's expression while an image, a video, or an ad is presented onscreen. Such technology would enable Facebook to improve content and advertising within users' newsfeeds. But this tool and

the data it collects will need to be used in ways that stay within the bounds of social norms.

Organizational training company FranklinCovey, which provides services mainly to businesses who seek to help their employees improve their performance, implemented an elegant approach to data use. The company leverages extremely personal data in a way that's an improvement on the social norm. Blaine Carter, data protection officer at FranklinCovey, told me that their self-assessment protocol includes requesting an assessment of employees by their colleagues. After the employee and her colleagues complete their assessments, rather than tell the employee under review "your boss and coworkers think you need to work on communication skills" (the social norm), FranklinCovey shows the person her results compared to an industry benchmark gleaned over the upwards of 30 million profiles FranklinCovey has collected, and the results are paired with a course recommendation. In other words, for the user, the results appear to be an objective measure compared to the average, rather than a judgment made by their coworkers. This improvement on the social norm respects the user's privacy and honors the person's ego, even when delivering a message that, in real life, could be uncomfortable, embarrassing, or insulting.

What Do Customers Want?

No company wants to jeopardize customer relationships or undermine trust by overstepping social norms. Yet sometimes understanding what consumers will deem appropriate or inappropriate is difficult. As a recent Pew Research Center study shows, customers themselves seem a bit confused. Researchers spoke to a focus group in depth about the data that was being collected about them, revealing an interesting gap between what people say they are comfortable with and what people choose to participate in when it comes to sharing data. For example, when asked if they would use a social media platform that allowed them to connect easily with classmates for a class reunion, but in return the site would use their data to target them for advertising,

51 percent said they wouldn't.[11] Yet, in another report, Pew said that 68 percent of adult Americans use Facebook.[12] Similarly, 32 percent of the focus group said they wouldn't participate in a loyalty card program with the condition that the company would sell their data to third parties. Yet American consumers hold an average of 13.4 memberships in loyalty programs, and they're active in about half of them.[13] How can people say they don't want to be targeted, yet hold in their wallets a dozen cards that allow companies to harvest their data? The two immediate explanations come to mind: *value*—customers will sign up if they receive significant benefits, and *lack of awareness*—many customers simply don't know how much data is collected about them.

However, the most significant factor is actually related to trust. Brands that have established trust with their customers get carte blanche. These could be leading global brands, or a niche brand that serves a particular audience of limited size. Customers decide whether or not to trust a company based on the perceived quality of the brand and the relationship they have with it. In many consumers' minds, trust looks like this: "I like this brand, and so far, nothing bad has happened. I see the value I get by giving these permissions to this company, so I will continue to do so." Consumers say they don't trust companies. But in their actions, I, along with industry experts, see something different: they do trust some companies, and they don't trust others.

At the same time that consumers are complaining about data privacy, they're also complaining that their shopping experiences aren't customized enough. Once again, this consumer desire for customization requires data collection. Ultimately consumers want to interact with the companies they trust, and they want those companies to know who they are, to a point, and on their terms.

Endnotes

1. Joseph Turow. *The Aisles Have Eyes: How Retailers Track Your Shopping, Strip Your Privacy, and Define Your Power.* Yale University Press, 2017.
2. `https://www.recode.net/2016/12/5/13842892/amazon-go-grocery-store-no-lines-cashier-paying`

3. http://www.natlawreview.com/article/white-house-issues-report-big-data-and-differential-pricing
4. Ian Glazer (vice president, Identity Product Management at Salesforce), in EIC Conference, May 2017.
5. Assaf Mischari (security expert, head of research at Team8, Former CTO in the cyber division of the IDF's Technology and Intelligence Unit 8200) in discussion with the author, May 2017.
6. Fatemeh Khatibloo (principle analyst, Forrester) in a presentation at Identified conference, June 2017.
7. https://www.forbes.com/sites/kashmirhill/2014/01/22/officemax-blames-data-broker-for-daughter-killed-in-car-crash-letter
8. https://dxc.turtl.co/story/55ee93d8bbfd077f2d4e22ee
9. https://dxc.turtl.co/story/55ee93d8bbfd077f2d4e22ee
10. https://www.thesun.co.uk/tech/3537504/pizza-advert-caught-stealthily-tracking-passers-using-creepy-facial-recognition-technology
11. http://www.pewinternet.org/2016/01/14/privacy-and-information-sharing
12. http://www.pewinternet.org/2016/11/11/social-media-update-2016/pi_2016-11-11_social-media-update_0-02
13. http://info.bondbrandloyalty.com/hubfs/Resources/2016_Bond_Loyalty_Report_Executive_Summary_US_Launch_Edition.pdf

6

The Search for a Better Self

This struggle between recording, protecting, and exposing information will continue, in ways that are hard to predict as everyone's digital identities evolve. This tension is the most interesting cultural phenomena that society is witnessing around digital identity. What lies deeper than the question, "What do people want to expose to organizations?" is the question "What are they willing to know about themselves?" Are people willing to see who they really are, as reflected by their data? Before everyone digitized their lives, it was easier to feel content about their choices and even moral values. Their actions weren't as readily recorded, if they were recorded at all, much less stored or accessible for others to review. But as more data is recorded and stored, people will encounter their true selves—not just who they say they are, but how they really act in the world. Are they really ready for that?

Learning About Yourself

Dave Asprey is recognized as the first person to sell something over the Internet. In 1993, before web browsers were widely used, he sold T-shirts that said, "Caffeine: My Drug of Choice." What's less known is that, around the same time, Asprey was struggling with a mysterious health issue. He was overweight, and despite all of his efforts to work out and eat fewer calories, he continued gaining weight. Then, at

the age of twenty-six, Asprey began experiencing memory loss. In an effort to measure brain performance, he used a fairly primitive game, Freecell, which established a performance baseline and proved that his brain was dull on certain days. "So now I'm fat," he said, "but I'm also getting stupid." Long before the "Quantified Self" movement got its name, with just this primitive means of measurement, Asprey became curious enough to record more data about his body and, essentially, his identity.

Although he couldn't discern the cause of his health problems, he continued to measure. By the age of twenty-eight, his company went bankrupt, and his stress levels increased so much that by age twenty-nine he was at increased risk of stroke and myocardial infarction. Brain scans showed he had low metabolic activity in many parts of his brain, but doctors didn't know how to address the problem, so he took on mastering his own body.

Through careful measurement and biohacking, Asprey was able to bring himself back to top health. Data was the key to his recovery. As much as he had tried to make the right diet choices and to exercise regularly, unknowingly he had been making the wrong decisions for his body. "You suck at knowing what your body is doing at any one time, unless you have specifically developed a skill to do that," he said. He developed systems that not only measured his body, but also correlated those measurements to his behaviors and his environment over time.[1]

By recording a wider range of his data and analyzing it, Asprey was able to assess more accurately the choices that improved the quality of his life. As he was doing that more often, a digital extension of his identity began to emerge.

Similarly, when astrophysicist Larry Smarr, PhD, found himself up against the limitations of medical science, he turned to personal data tracking. Smarr was diagnosed with an inflammation in his colon. His doctor told him the nature of the problem, but Smarr wanted to know the source. "I'm a little obsessive, so I've been taking blood tests for the last five years as well as stool tests," said Smarr in his TEDMED

talk. He tracked approximately 150 variables in his body, and he found that only one of the measures was out of the norm, something called CRP. But when he explored further, he learned that hundreds of other microorganisms could be identified. "Stool is this wonderful window that is completely ignored. It's the most information-rich medium you've ever laid eyes on. Half of it is microbes, about a billion per cubic centimeter, and each of them are about 10 million bases of DNA," he explains. However, Smarr was unable to get his doctor interested in the results of the tests he had done, and his doctor was unwilling to look at these microorganisms. "If you think coming in with a pile of printouts from the web is an informed patient, I'm like the doctor's worst nightmare," he joked.

Smarr discovered that, despite the detailed information that can be measured, doctors don't look at this type of data, nor does science even know what all of these different microbes do. Smarr found a doctor who was willing to work with him and look through his patterns, comparing them with those of healthy people. They've since worked together to make changes that are helping Smarr heal his colon.[2]

Smarr is only one of thousands of people who recognize that medical protocols should be personalized to an individual's specific body, history, and condition, and that the medical profession isn't equipped at this point to track, manage, and analyze all this information for every patient. That's why these people are taking their health into their own hands.

Today, many people are trying to improve their lives—their health, finances, stress levels, and overall happiness—by transferring personal data into the digital world in exchange for analysis, feedback, and advice. As this trend grows, it raises important questions about human behavior and identity. How is this self-focused data changing people's identities, making them faster, stronger, smarter, richer, and maybe even more ethical? What kinds of decisions does this data lead people to make? In other words, how can your identity data shape your actual identity?

How Do You Decide?

It could be said that you're essentially the sum of every decision you've ever made. In his Nobel Prize–winning work that examines decision-making, Daniel Kahneman stipulates that people use two methods:

- **Fast thinking:** *Fast thinking* is necessary for intuitive behaviors and habitual decisions, such as driving a car and reading other people's facial expressions.
- **Slow thinking:** *Slow thinking* is used to evaluate non-urgent decisions, like if it's time to look for a new job or what kind of car you want to buy.

Often, people use fast thinking when slow thinking would be more appropriate. The most obvious example? What you eat. People tend to eat habitually, or they eat what their peers or culture demand. Few people take the time to analyze how their diet is affecting them or how changing their habits might improve their health. Yet, it's this kind of slow thinking that enabled Asprey to improve his bodily health and brain function dramatically. Rather than accepting common knowledge or eating what he was accustomed to eating, Asprey objectively measured outcomes and made calculated nutrition decisions, which led him to heal himself by eating a diet that contains 70 percent of calories from fat.

Perhaps you've noticed more restaurants listing the calorie count next to items on their menus or grocery stores noting the sources of the foods that they offer? This information invites you to slow your decision-making processes and to make your food choices based not just on instinctual cravings, but on deeper personal values concerning your health or your commitment to buying locally sourced foods. Similarly, the data you collect about yourself can help you make slow-thinking decisions in your fast-moving life.

The Three Considerations of Personal Data

The collection of personal data is pervasive today, whether it's measuring your weight, the number of footsteps you take, the temperature of your environment, how much and how well you sleep, who you communicate with, your health records, and so on. Dan Ariely, founder of The Center for Advanced Hindsight and one of the leaders in studying how humans make decisions, says that as this data is collected, people should consider three questions: "What do you want to know?," "What do you want to share?," and "What impact can this data make?"[3]

What do you want to know?

According to Ariely, what you want to know depends upon your circumstances at any given time. For example, if you have no intention of changing your spending habits, looking at data on your spending isn't much use, and might even upset you. You don't want to know your heart rate most of the time, but if you're exercising or if something goes wrong, you definitely do. You want to check your bank balance and cholesterol levels periodically, but you don't want that information all the time.

Ariely points out that knowledge isn't always helpful. For example, in certain scenarios, some doctors recommend against telling a patient he has an incurable disease. Furthermore, sometimes data-driven choices have a downside. For example, consider end-of-life decisions. In the United States, a family chooses when to take a patient off life support, but in France, the doctor makes this choice. Studies show that, in the US, no matter what a family chooses, they experience more distress than the families who didn't have any choice in the matter. For American families, having the information and the power to choose hurts more than it helps.[4]

Figure 6-1: *Banks realize customers sometimes prefer not to view their balance when they withdraw cash.*

Some companies that collect personal data understand this point. It's not surprising that, when I received my DNA results from 23andme, before exposing information about potential disease findings, they asked me—twice—to confirm that I really, truly did want to know if I may be at risk for a possibly incurable disease. They understand that having such data can sometimes do more harm than good. (I clicked yes, and I'm fine, thank you.)

What do you want to share?

Sharing information raises questions of both trust and utility, Ariely says. People share intimate personal information with others in their lives, as part of relationships. People share information with healthcare professionals to manage their health, and they share yet more information with companies who sell them products and services. When deciding what to share, you need to ask yourself, "Do I trust this entity?" If you do, then ask, "How will sharing this data benefit me?"

What impact can this data make?

Access to data can help people make calculated, slow-thinking decisions. Making better decisions might be the difference between life or death, or it simply might improve the quality of your life.

Introducing the Quantified Self

Several communities are sprouting up around the idea of making themselves better through data. One of the better-known groups, the Quantified Self (QS) movement, uses self-measurement of exercise, body function, foods, moods, environmental stress, behaviors, and more, in order to live longer and maximize the quality of their lives and their performance. Athletes want to improve their bodies, professionals want to perform better at work, and, well. . . everybody wants to improve their moods and behaviors. By measuring everything from footsteps to stress levels, people are looking for happiness, growth, and overall improvement in their lives. An entire industry has grown up around this desire for data-driven self-improvement.

The success of older athletes such as Tom Brady and Dana Torres can be attributed, at least in part, to the level of monitoring now available. These athletes have continued to deliver peak performances, even as they have aged, challenging traditional notions of what can be achieved in professional sports, beyond the age of thirty or even forty. Now a technology startup named Whoop pays top athletes to wear wristbands that track their body signals a hundred times per second, to study their routines and their recovery, and then to sell that data to people who wish to improve their performance like top athletes do.

While writing this chapter, I began analyzing the different types of exhaustion I experience as a runner. By measuring my sleep, nutrition, temperature, and a few other basic parameters, I was able to identify the subtleties of different types of exhaustion. Now, if I feel weak towards the end of my 10K I can better identify the sensation—be it fatigue, lack of glucose in my muscles, unrecovered muscles, or the temperature outside—and make adjustments accordingly.

Knowing yourself, inside and out

For most people, the idea of measuring themselves constantly is just a bit obsessive, partly because they think they're making good decisions already. But as Kahneman's work shows, people make a variety of errors when evaluating past decisions. First of all, to assess their decision-making, humans need a coherent picture. "We are evidently ready from birth to have impressions of causality, which do not depend on reasoning about patterns of causation," he says.[5] People also tend to ignore evidence that doesn't fit in with their beliefs. In other words, your brain, particularly in fast-thinking mode, has myriad ways of fooling you.

During a training in which he attached diodes to his body to measure brain function, Asprey saw firsthand evidence of this. He wanted to learn to detect when he was lying, not only to others but, more importantly, to himself. "I learned that I have phenomenal powers of self-deception,"[6] he says. The desire for improvement is eternal, but the desire for people to think of themselves as good people is even stronger.

Even when people track their data, they may prefer some measurements stay hidden from them either because they aren't motivated enough to pay attention, because making a change is hard or because they just don't want to know the truth. Face it, if all human beings wanted to be healthy, they wouldn't use intoxicants or smoke cigarettes or consume sugar. According to Ariely, when knowledge puts people in a dilemma, they prefer not to know.

The technology

Even for people who want to use more data in their lives, recording and accessing the data isn't particularly convenient. Take a smart scale that records your weight but doesn't display it. In order to see your weight or to see charts showing your weight fluctuations over time, you'd have to open an app on your phone. Today, that's how the interface for most tracked data works. You need a special device or app to record it, you need to turn the device on, and then you have to log in regularly to view it.

The clumsiness of this process is still the main challenge. The technology industry can develop more immediate ways to let people know if they aren't drinking enough water, whether electromagnetic fields are above their personal threshold or how far behind they are on their daily pedometer goal.

Having an internal device automatically record your measurements would take much of the hassle out of collecting data. Some biohackers are experimenting with such technology, implanting chips in their bodies to track vital data. But most people aren't interested in such extreme forms of measurement. However, devices such as heartbeat and movement wristbands or glucose monitors are now common. Some companies are deep into development of headphones that measure things such as pulse and blood pressure.[7] It's just a matter of time before companies figure out easier ways for people to measure physical activity without having to go to the trouble of, well, measuring it.

The future

Today, the people who conscientiously and consistently track themselves are the outliers, but the way they are using data to change themselves is an important indicator of how everyone's futures may look. As data collection becomes simpler, it's worth thinking about what might get collected. Someone who snaps one picture each day may want simply to record a memory of her life. But as this *lifelogging* trend grows, what might people learn about themselves from frequently captured visual records?

Someone who purchases a wristband to measure her footsteps may want to collect only limited data, but as the technology advances, much more information will be accessible. Theoretically, if you were wearing a wristband that measures the fine details, sweat, and other emotion-driven bodily responses, you could track your emotional responses to people by noticing heart rate increases that correspond to dialing a particular person's phone number or seeing a certain name in a text message. Your device could know if you have a crush on someone or if a person causes you stress before you realize it yourself.

You instinctively know certain situations and people stress you out, but when you can combine your personal data, such as mood and heart rate, with environmental data, such as the people or noise levels around you, you can gain more objective insight into the factors that raise your stress levels, and you can learn how to mitigate that stress. As data collection capabilities develop, what insights do you think you might want?

What Can Be Measured? What Can Be Improved?

Here I discuss the different aspects of self-measurement. I began the chapter with two stories about personalized health treatment because that's the most natural starting point. Medical professionals often speak about the practice of medicine as statistical analysis. Based on certain symptoms, doctors make a statistical diagnosis, identify the probable cause of an illness or disorder, and then treat the patient accordingly. If the treatment doesn't work, that information feeds back to the doctor, who factors it into the next diagnosis. Currently, all of this diagnostic and treatment information is generalized. In other words, the same statistical analysis is performed on each person, without any regard for each person's specific body and genetics.

How will medical practices change when those statistics are personalized? Already, nutritionists agree that different people's bodies respond differently to diet—there is no one-size-fits-all health regimen. Personalization, then, is the next step in modern medicine.

Originally, when scientists spoke about personalized medicine, they looked to DNA sequencing for additional insight into patient profiles. But now, with so much personal data being collected, doctors can analyze countless statistics and patterns—everything from proximity to a power plant to nightly sleep patterns. Ideally, these capabilities will help doctors both diagnose patients and predict what therapies will work best for each individual.

The amount of research required for personalized medicine might seem overwhelming, but as more data about more people is collected, much of the research is already built in. Personalization becomes less about advanced research techniques and more about big data analytics. Sites like patientslikeme.com allow individuals to compare their pathologies and treatments to one another and to identify patterns that they can adopt to help themselves. In collaboration with Data for Good, patientslikeme.com not only helps patients who use its platform, but also donates collected information for research purposes. As more data is collected, it's only a matter of time before big data analysis allows many correlations between so many more from the altitude at which someone lives to their dietary habits. Then doctors can personalize medical treatments and drug dosage based on statistics relevant to a patient's individualized profile instead of statistics generalized over an entire population.

Can data make you happy?

Clearly, data-based decision-making can help you improve your health, but could measuring yourself also lead you to greater happiness? Not just the "Hey, eating that slice of cake would make me happy for a few minutes" kind of happiness, but a deeper, lasting satisfaction? According to Kahneman, it might.

Overcoming cognitive biases, in some aspects of life, may be the key. For example, Kahneman suggests it's difficult for humans to evaluate how much pleasure they derive from particular activities.[8] Why? We have two selves, an "experiencing self" and a "remembering self." For example, a person who has had a great experience on a vacation that ended with a horrible last day tends to remember the entire experience as terrible. Conversely, if this person had a poor experience that ended wonderfully, he tends to recall the entire experience fondly.

In addition to contradictions between the experiencing self and the remembering self, people often have difficulty correlating their experiences and emotions. For example, if a coworker says something that irritates you, you may correlate your irritation to that coworker's

comment. But often, an irritating comment isn't the primary cause of your irritation—it's just the final straw. Maybe lack of sleep, hunger, background noise, electromagnetic waves, low blood sugar, even pollution, was causing irritability long before your coworker made that comment.

Without accurate measurement of your emotions and your surrounding circumstances, it's really difficult for slow thinking to occur. It's not that people are lazy, Kahneman stipulates, but that they're just too overwhelmed by incoming information to enact the slow thinking they need to take the best possible care of themselves—to make the choice that leads to greater happiness—in any given moment.

How can self-measurement help you with this? It's as simple as correlating cause and effect. The challenge is finding out what measures are relevant to the results you seek. Several currently available apps measure your state of mind. The simplest apps ask you to fill in basic information, like about sleep and nutrition, and at random intervals they prompt you to click on a mood icon or give a numerical rating to represent your mood. More sophisticated solutions can measure biological indicators and even brainwaves. Understanding how your moods and your hormones affect your choices can be the first step toward making decisions that lead you to the happiness you seek. At the moment, only biohackers are addressing the challenge of correlating their moods with other data, but as technology progresses, access to such information will be available to more people.

Happiness, balance, personal fulfillment—achieving these states requires much deeper analysis and attention than these solutions can provide. Yet the tools and methodologies reviewed here, although only beginning to help you understand who you are, offer an interesting approach to self-exploration and improvement.

Measuring lifestyle

Ahnjili ZhuParris, a student in The Netherlands, wanted to see what she could learn by measuring her behaviors during her leisure time. She wanted to improve her productivity in college and to cut back on the time she wasted browsing the Internet. While she was mea-

suring her Internet use, she was also measuring her menstrual cycle to find out if her hormonal changes affected her behavior. They did. "There was no change in the amount of money I spent shopping or the amount of time I spent shopping online, which was great to learn. . . but I did find something quite strange. If I was shopping at the time I was fertile, I actually bought more red-colored items. In fact, it was the only time I bought red-colored items."[9] ZhuParris also found that her cycle influenced her Tinder swipes. When she was most fertile, she swiped right more often on men whose pictures were taken at the gym and, generally speaking, she was less selective in her choices.

Since analyzing this data, ZhuParris has reported that her awareness has changed. She's now able to step back and ask herself about specific behaviors and how her hormones might influence them. Meanwhile, she's moved on to a new area of study. "I've been experimenting with psychedelics, both the inputs and outcomes. So, I've been looking at micro-dosing, and also in taking full doses, and which factors shift my psychedelic experience," she says.

While using data to optimize a person's health or to save people's lives may sound like noble pursuits, sometimes data can be used simply to improve the quality of people's lives—to have more fun, to date people who are more compatible, or to make better use of leisure time.

Challenges

In today's western culture, everyone wants to be more something—more fit, more efficient, more productive, more affluent, more disciplined. . . . But, when self-improvement requires confronting the truth, many people follow the old adage, "ignorance is bliss." What they want, at the end of the day, is to feel good about themselves with minimal effort.

In the realm of fitness, pace trackers are one of the most popular quantified self apps for a reason: everyone is already walking. It's easy to park farther from the shopping center, walk across the parking lot, and call it an accomplishment, so people do it.

For those who want to feel environmentally responsible, similar quick fixes are available, like apps that allow people to compare the carbon emissions between train, car, and airplane travel. Even if you make the better choice only half the time, you've made an improvement, without making any big personal compromises.

But, if an app reveals a truth you're not ready to hear—like a budget app that suggests your spending habits don't match up with your savings goals—you're likely to stop using it.

Ultimately the most successful self-quantifying apps and self-improvement hacks are those that give people a quick fix. They collect data silently, aggregate it, and display it to people at a time and in a way that makes decisions easy.

As developers think about what apps people will actually use, this easy-fix aspect is essential. People are willing to make some changes, but most people won't make major changes quickly, nor will they make changes that are complex or inconvenient to them.

A further difficulty with current data collection techniques is that each app collects its own data, so anyone who wants to figure out correlations between different factors needs to find a way to collate the data, either in spreadsheets or by writing code. As one man discovered, often people are surprised to learn the causes that correspond with their results. To prevent himself from needing to take bathroom breaks during important meetings, he measured his water-drinking, only to find out that bathroom breaks were inversely correlated to how well he prepared for the meeting, not to his fluid intake.

This area has infinite business potential—creating applications that help people understand the correlations between different areas of their lives, so they can make better decisions. As Smarr discovered, there are thousands of data points that medical science doesn't even track, much less analyze. To do so would require tremendous computing power as well as lower cost and simpler solutions for data collection and reporting. Few people are willing to do a weekly stool analysis, as Smarr was, and even fewer are willing to track a hundred thousand microorganisms, even when their lives depend on it. Again, when it comes to both medical science and behavior modification,

the business challenge is to simplify the analysis and to provide easy decision-making criteria that an average user can understand.

Virtual-World Data, Real-World You

Today's biohackers and quantified-self members seem like the outliers, wearing the latest trackers, obsessively checking their vital signs, eradicating Wi-Fi and Bluetooth from their bedrooms, and measuring the influence of adding butter to their coffee. But, you might not be so different from these folks. If you own a smartphone—and it's safe to assume most everyone reading this book does—you too are tracking your behavior, at the very least your location and communication data. Perhaps you also track data in a financial planning app or one that supports your workout routine or meditation practice? What about steps taken per day, receipts for business expenses, likes on social media, or the temperature in your hometown? In fact, every photo you take is another data point you record about yourself. If the story of your life were based on this data, what would the plotline be? If you tracked every decision you make, what would the data reveal about you?

But some people are tracking themselves even more than biohackers. In the virtual worlds gamers inhabit, they constantly amass all kinds of personal data—often without intention—from the people with whom they associate to the places they visit, from what they bring with them to what they build or destroy, from how many hours a day they play to how often they get up for breaks to how late at night they play. Although these games are make-believe, gamers' data on these platforms can be quite revealing.

Every move on a multiplayer game is recorded, and multiplayer games are becoming big business. For example, League of Legends had more than 14.7 million simultaneous live viewers for the 2016 world championships with winners earning $6.3 million. Players for these games have become professional or semiprofessional competitors, and all of their stats are available, just like stats for any professional athlete.

So, what does gaming life have to do with real life? Your behavior when you're playing a game is your personal behavior, and your game data is your personal data. It says a lot about your problem-solving skills, your leadership skills, your reactions to adversity, your commitments, and many times, your values. Virtual worlds can show you—and others—who you are without a filter.

Your online identities, even if they're not attached to your real names, become so real that friendships that begin in the virtual world can follow you into the real world. Long-time players in particular game worlds are often celebrated, helped by the community, or mourned upon their death, in the real world. For instance, when a father in the World of Warcraft community lost his son, the community supported him during his mourning.[10] Although players may never go as far as meeting each other in real life, they are part of a community of real people, and they are respected for their accomplishments and their identities, as expressed through their actions and behaviors, within that community.

You can look at game identities in two ways: someone's in-game behavior is just one more set of data about that person, or you can see it as making up one of many distinct personas they may have created. In your personal life, you're probably aware that you have different personas. You act differently at home than you do at work. You may be a leader in your spiritual community, but you're the last of the pack in your weekend cycling group. In virtual worlds, you consciously create alternate personas, which may reflect parts of yourself that you never express in the real world. But, unlike the real world, here your every move is recorded and stored.

Does this notion change the way you're willing to behave in the virtual world? How would recording all your real-world decisions change who you are in real life?

Remembering Who You Are

What would happen if you had a traumatic accident that resulted in total amnesia? If, when you woke up, you had to relearn everything?

Would you still be the same person? Most people say no. For the vast majority of everyone, this question will remain hypothetical, but in some ways it does reflect reality. People's memories are faulty, as are their decisions. People tend to remember only selective information about themselves and tend to see themselves as better than they are. But, as you record your information and receive better guidance on which to base your decisions, increasingly you'll be forced to face the truth about yourself. What happens then?

According to Kahneman, people make a variety of errors of association, as they always have. For example, your immediate environment can influence your thoughts, and your thoughts, of course, influence your behavior. Did you know that someone who thinks about elderly people will automatically walk slower? Additionally, people are swayed by confirmation bias. When they begin to notice some patterns, they ignore evidence to the contrary. They're further influenced by substitution, representativeness, or leaps in judgment, drawing conclusions based on too little data. Also, as Kahneman notes, everyone tends to be biased toward things and people they like.[11]

Because careful decision-making requires people to remember all the relevant facts and to put aside all their biases, it takes tremendous effort. More often than not, people are unsure of which decisions really benefit them, largely because they don't allow themselves to see the evidence. For instance, you may have a sense that you're unhappy in a job or in a relationship. Yet every day you find reasons to stay, ignoring clear evidence that suggests it's time to move on. It's only when you finally make the decision to leave that you look back, see the evidence, and wonder why you stayed so long. Perhaps this denial is human nature—many people seem hardwired to resist big changes.

As people track more data, they may no longer have such selective memories. A variety of devices will record their decisions and their outcomes, allowing them to make more intelligent and calculated choices. If you're wearing a wristband that records your sleep, you'll know within a matter of days whether you always drift off at the same time in the afternoon, whether you drift off a specific number of minutes after having eaten lunch, or whether it happens only if you've eaten a certain kind of food. What Kahneman refers to as "slow

memory" will become more automatic, and recorded data will help you make decisions based on true, long-term trends, instead of your faulty memories and biases.

Imagine the far-reaching possibilities. As these apps are used in education, data may help you make better career decisions. If you're measuring your brain function and energy levels, you can make better decisions about when to perform the most challenging tasks at work and when to schedule more mundane tasks. You'll know to hold meetings when you're most attentive (or least attentive, if the speaker is boring).

While working in alignment with your data can help you perform better, it also raises questions about who you really are. Your destiny is determined in large part by the decisions you make. Using data as the basis for your decision-making will enable more slow thinking and, in fact, make slow thinking faster, because you have better access to data from past decisions. But could the shift from fast thinking decisions to slow thinking change people as human beings? As you better understand the impact of your decisions, will you actually make better choices? Will you make more data-driven and fewer emotional decisions? Will you lose faith in your own intuition and therefore use it less?

Making Better Decisions

Kahneman's work suggests that the biggest barrier to thinking more logically is the effort involved in recalling and analyzing the true information about any given situation. Recording personal data will simplify recall, giving people access to their experiences and wisdom at a glance, thereby reducing that barrier. In this way, personal data collection will present everyone with real opportunities to make better decisions and to improve many aspects of their lives.

We as a society now have unprecedented abilities to measure factors that are important to us and to set our own standards for the areas we want to improve, whether it's our health, productivity, happiness, or the way we engage with and treat others or the environment.

Already we see people at the cutting edge of the quantified self and biohacking trends, reaping benefits, recovering faster from disease, experiencing higher energy levels, sleeping better, and getting greater enjoyment out of both their work and leisure time. As these technologies become pervasive, people can expect to face a dilemma: How much of myself am I willing give up in order to reap these benefits? At this point, people will still be in control, because they will choose what they measure and share. But what happens when they are asked to give up agency as well? I explore that in the next chapter.

Endnotes

1. https://www.youtube.com/watch?v=h8hpjwKESXw
2. http://www.tedmed.com/talks/show?id=18018
3. Dan Ariely (professor of psychology and behavioral economics, founder of The Center for Advanced Hindsight) in a discussion with the author, March 2017.
4. http://magazine.columbia.edu/features/winter-2010-11/grave-decisions?page=0,1
5. Daniel Kahneman. *Thinking, Fast and Slow.* Farrar Straus and Giroux, 2011.
6. https://www.youtube.com/watch?v=h8hpjwKESXw
7. https://arstechnica.com/gadgets/2016/12/hear-the-pulse-heart-rate-monitoring-fitness-earbuds-tested/
8. Kahneman. *Thinking, Fast and Slow.*
9. http://quantifiedself.com/?s=Ahnjili+ZhuParris&x=0&y=0
10. http://kotaku.com/dad-says-hes-mourning-his-son-wow-community-rushes-to-1682284173
11. Kahneman. *Thinking, Fast and Slow.*

7

Identity and Artificial Intelligence

Lately my Gmail account has been suggesting short responses to emails I receive, things like "Great!," "See you then!," and "Sounds good to me!" When have I ever said "Sounds good to me!"? When have I ever used an exclamation point, for that matter? Yet, suddenly all kinds of things are sounding good to me, exclamation point! I hope my friends are enjoying my new, bubblier version. But seriously, I wonder what will happen if this system gets just a little bit smarter and starts offering responses that actually sound like something I would say? Or what happens if it doesn't, and I start using these pre-programmed responses, because it's so much easier than composing my own? Will I start to create a new me who is a lot more upbeat, at least in text communications? Maybe that new me will be a lot more like the new you?

As you and I and everyone else amasses more personal data in the digital world and as technology uses this data to make our lives easier, how will our identities change?

Who Is Making Your Decisions?

A big part of your identity is comprised by the decisions you make. But are you really making these decisions on your own? Tech entrepreneur and investor Naval Ravikant references a theory that the emotional

and social behavior of a chimp is the average of the five chimps closest to it. In this way, humans aren't so different from these evolutionary cousins. The people you associate with have a deep influence on your identity, whether you like it or not.[1]

Some people you enlist to influence your identity. For example, you may seek a particular friend's advice, because you see that person embodying values you respect. Or you may choose to guide your decisions based on what you believe a mentor or teacher you admire would do in a similar situation. On the other hand, some people's influence on your identity happens without your awareness. Your family, for example, shapes your identity not only genetically. Have you ever found yourself mindlessly doing something that your parent used to do? Or has a phrase your mother used to say suddenly fallen out of your mouth? In these ways you unwittingly adopt aspects of your identity from the people around you.

In the digital age, computing devices are becoming part of people's identities in the same ways that other people do. When you ask an app for a recommendation for a hotel, you put the app in the role of trusted advisor. Some people might choose TripAdvisor, while others might use Yelp or Google for their mentor, and the mentor whether through aggregated content generated by people or other means, will shape how they plan their trips. Similarly, when you choose an app to help you invest your money, one app will advise different spending, saving, and investing patterns than another. These digital advisors become part of your decision-making processes, influencing how you think and who you become.

Taking this further, people configure their devices to match their preferences, so they can start to think of a specific device as a part of them. If you left home without your mobile phone, more than likely you'd go back to get it, because you feel you can't function in the world the way you usually do without your phone. Similarly, when you speak to your Alexa or HomePod, the device begins to take on an aspect of your personality. As it provides more and more advice to you, such as how to dress for a particular occasion, what music you might like, or what health habits to adopt, it becomes more than a device or advisor. It becomes an extension of your identity.

As I discussed in the previous chapter, people can use data to help them make decisions that result in happier and healthier lives. But how much agency are you giving up in the process? How much are you letting your apps and devices influence your identity?

What Isn't Known

Fundamentally, people can utilize two categories of data: data that is collected about themselves personally and data that is aggregated over large populations. Personally collected data can give better answers to questions such as "What should I eat before a workout?" Aggregated data from multiple sources can give answers to questions like, "What do other people eat before a workout that I haven't tried yet but that resulted in maximum performance?" Even more specifically, data can give answers to questions like "What do white male runners between ages forty to fifty eat before a ten kilometer run in hot weather?"

Furthermore, computing systems are gaining the ability to offer you answers to questions that you don't even know to ask and to find correlations between a multitude of factors that create patterns that are so complex that they're not easily perceptible or understandable to the human brain. Yet a computer can make recommendations based on this data.

So, if I want to decide what to eat before or after my next run, computer analysis systems can look at multiple relevant factors—the things I can think of that will affect my run, such as my recent sleep patterns, workout routines, and the weather, as well as factors I wouldn't think of myself.

Data-based decision-making happens on two levels. Take a medical diagnosis as an example:

- One must decide **what data is relevant.** Generally speaking, from the tracked data you don't know how doctors decide what is relevant. Are they making comparisons and presumptions based on your demographics, your behavior, your body type, or your associates? Generally you don't decide on the

parameters of comparison that lead to a doctor's diagnosis or recommendations.

- One must decide on **a course of action.** Based on the gathered data, doctors give a diagnosis, and they might provide options for treatment, perhaps with a success rate for each. You get the final decision about your course of action.

These same levels of decision-making happen with all data-based decisions, whether I'm talking about an app that helps navigate my hometown or a social media site selecting content for my newsfeed.

People don't often think about the factors or calculations their devices use to formulate their recommendations; they simply input data and receive responses. Some of the means by which they interact with their devices call attention to this exchange—they type data into an app, or they tap a screen. But other interfaces are becoming seamless, like a voice interface that tells their home device to play a certain kind of music or a wristband that informs them when their insulin level falls below a certain point. They don't think too much about how the device decides what songs to play—or if it's limiting or expanding their musical horizons—they're just happy to be listening to music they like.

The more seamless the interface, the more a device comes to feel like part of you. If my home device knows the temperature I like when I arrive home from work, the lighting level I want at dinnertime, the music I want to hear before I go to bed, and the time I want my coffee to start brewing in the morning, the device begins to feel not like an outside influence on my decisions, but more like an extension of my identity.

Artificial Intelligence Enters the Picture

Your entertainment choices or ambiance settings aren't particularly important decisions. But when computers are used to make critical decisions that directly affect your well-being, the success of your

enterprises, and the day-to-day management of your life, things get really interesting. What once was merely fodder for science fiction is becoming a reality: artificial intelligence.

Artificial intelligence (AI) is a set of techniques that aims to enable a computer to mimic the decision-making process of a human brain. As AI technology progresses, there are an increasing number of AI-supported applications across many areas. The beauty of this computerized decision-making is its *scalability*, which, for example, enables it to provide deep analysis on an individual level.

One interesting development is taking place in the field of agriculture, where AI is being used to make personalized decisions down to the level of a particular plant. In a recent project, IBM and E. & J. Gallo Winery collaborated to dramatically increase the yields and quality of grapes by connecting microsensors to each individual plant and using technology to deliver precisely the amount of water and fertilizer each plant needed each day. The solution also utilized satellite data for tracking weather patterns and soil types. In this system, analysis goes as far as considering the leaf canopies of each plant as well as the grape-producing history of each plant. Just like humans, each plant is an individual, and using a combination of aggregated data about all the individuals plus data about each individual, this experiment showed that both the yield and quality of the harvest could be improved by 25 percent by delivering individualized nourishment to each plant.[2]

What AI did for plants may suggest what humans expect in the future. Until recently, many people thought of AI as addressing a narrow set of domains of decision-making. Playing chess, for example, is a specific domain, with fixed rules and fixed logic. Yet teaching a computer to beat a human chess championship took years of programming. Developers hired chess grandmaster Joel Benjamin, programming his knowledge into Big Blue. The example of plant health takes a different approach to computer intelligence. Sensors at a microlevel measure multiple factors on and around the plant, including perhaps the height of that plant and the height of the one near it. All of that data is recorded over time to understand the trend of that individual as well as the other individual plants and their regimens.

Moving out, the system needs to know the predicted weather. If it rains, not only does the plant need less water, it also might need more fertilizer because the fertilizer is washed away in the rain. Of course, if the plant is in the middle of the field, there will be more runoff from surrounding plants than if it's at the edge of the vineyard, particularly if it's uphill rather than downhill. And so on.

The number of factors that can be calculated in this project is overwhelming. As IBM develops this technology, the value is clear—how such a system could make extremely complex health calculations that humans could only guess at. For example, everybody knows that people need to drink more water when the weather is hot, but how much more and what is considered hot? Someone accustomed to a hot environment might have a body that manages water loss more efficiently than someone who just flew in from northern Canada.

As I discussed in the previous chapter, human decision-making is subject to many faults, both because people make most of their decisions quickly or instinctively and because even when people take more time on their decisions, they're subject to a variety of biases. There are some high-impact decisions that we know we do a poor job at making for ourselves. For example, philosopher Alain de Botton famously argues that people are destined to marry the wrong person.[3] Would people delegate that decision to a computer? As historian and philosopher Yuval Noah Harari points out, perhaps people wouldn't fully delegate it. They would want the final word, but they'd want to know what the data says.[4] His example, in which a woman receives a recommendation to marry one of two suitors, doesn't represent how life really works. They wouldn't put together a list of the people who somehow simultaneously proposed marriage to most people and then create a decision table. They would generally make or receive one proposal at a time.

But what if a computer could analyze the probabilities of your happiness if you were to marry a particular person? Would you want a computer to tell you, say, that the probability of divorce is 82 percent, statistically speaking, but you'll still be happier if you marry this person, have two children, and then divorce, than if you didn't marry at all? Or to tell you that your chance of finding a spouse who is a

better fit is only 34 percent? It could be just like having an extra Jewish mother in your pocket, reminding you that your biological clock is ticking.

How Far Will AI Go?

So far I've discussed how computers do a good job calculating data, showing you your data, and making recommendations. But how far are you willing to trust your computers? How far can they go?

The nuances of human relationships are complex. Although an app might give you tips on what kind of date would be fun for you and a mate, many people probably wouldn't want it to tell them if they should continue dating someone or what they should say in a conversation. However, in his book *Homo Deus: A Brief History of Tomorrow,* Harari postulates that people's automated devices will begin making deeper decisions for them. He believes that two people using Cortana, the personal assistant being developed by Microsoft, could enable their assistants to exchange information with one another to determine if they are, for example, likely to enjoy a biking trip together. He goes so far as suggesting that, if you have Cortana managing enough of your life, someone with a better version of Cortana would have better job prospects because their version would know better how to "speak" to an employer's Cortana.[5]

How smart will AI become? And how soon? Some people predict that computers will become more intelligent than human beings by the year 2030 and that their intelligence could be a threat. Harari warns that if people want to understand how intelligent computers will treat humans in the future, they ought to take a look at how humans treat animals of inferior intelligence today. Others predict doom because a narrowly targeted AI solution will relentlessly pursue a goal, such as the comic example of handwriting perfection, as explored on the blog "Wait But Why," in which a fictional program called Turry decides that wiping out humankind is the most expedient way to accomplish this goal.[6]

On the other side, many scientists and AI experts argue that both of these scenarios are too far-fetched. AI probably won't replace human thinking. It's close to impossible for a computer to be truly intelligent because AI doesn't have common sense. There are so many rules of common sense that people never state explicitly, for example, what secrets are good to keep, like a surprise party, and what secrets are bad to keep, like that a government official offered you a bribe, which is why there are no computer whistle blowers. In both cases, you may have been told to keep the information secret. In both cases, the consequences of telling or keeping the secret are obvious to humans in any culture, but not to computers. Many experts predict that computers will never truly have beliefs, emotions, or values, nor will they feel pain or pleasure. What AI will do, they argue, isn't replace human brains, but augment human brains. "When you use a phone, you amplify the power of human speech. You cannot shout from New York to California," says computer scientist Sebastian Thrum, in an interview with *The New Yorker.* "Did the phone replace the human voice? No, the phone is an augmentation device. The cognitive revolution will allow computers to amplify the capacity of the human mind in the same manner."[7] In other words, AI will enhance, not replace human identities.

Your Agents

Computers, like your human advisor, can give you a variety of recommendations. As time goes on, they'll be able to offer increasingly more intelligent and complex recommendations. Eventually, they'll have the capability to make decisions for you, acting as your agent.

Just as you may feel perfectly comfortable having a personal assistant or family member representing you in specific situations, eventually a computer system will be able to represent you too. Take for example a computerized system that sets travel plans. Travel planning means making a series of decisions based on a defined and limited set of considerations. My computerized agent can know enough about me to do that. If the system my agent communicates with to make my

reservation says, "You can have a window seat on this flight, but if you want a window seat on the connecting flight, it will be another $55," then my agent will know enough about me to say yes, I'll pay, or no, I won't. For all intents and purposes, the computer is representing me to the other computer for this transaction.

In order for machines to represent me and make decisions for me, three things must happen:

- Machines must be able to understand accurately my requests and all the implications of them. At some point, I'll be able to say, "I have appointments in London on June 8 and appointments in San Francisco on June 10 and 11. Please make all the arrangements," and Alexa, Google, Siri, or Cortana will be able to understand that I need flights, hotels, taxis, and possibly some other arrangements.
- Machines must hold all necessary data about my identity and my preferences. The system needs to know I prefer day flights and window seats, my company credit card should be charged, and I'm willing to take a stopover if it saves a few hundred dollars and doesn't extend the duration of my travel by more than three hours.
- Interfaces must be able to speak to one another. As *bot* (a computerized agent that usually represents companies in interactions with consumers) technology develops, more companies offer such systematic interfaces. Today, Google can order an Uber but not a Lyft, because each integration is customized. Tomorrow, bots and application programming interfaces (APIs) will make it much easier for machines to interface one another, so your Alexa, Google, or Apple device will be able to communicate with airlines, cab companies, and hotels, and then optimize all of the details through automated interfaces.

Ultimately the decisions you delegate will depend upon the accuracy of the available solutions, weighted by the importance of the decision. For example, if I ask Alexa or Google Home to do my grocery

shopping, and it gets me the wrong breakfast cereal, the cost of that mistake is low. On the other hand, if my automated personal assistant books an itinerary with multiple stopovers, a noisy hotel room far from my meetings, or tickets beyond my company's approved budget, the consequences are higher.

Challenges with AI

As computers make more and more decisions, it's important to be aware of the challenges of AI, which I discuss here.

Computers don't have common sense

As mentioned previously, computers don't have common sense. Their matching algorithms are just matching algorithms, which explains why in 2014, when IBM's Watson was asked to name the first woman in space, the answer it came up with was "Wonder Woman."[8] Data scientists have since taught it to give the correct answer, but this example illustrates one angle of the problem. What's obvious to us as humans isn't obvious to a computer. We can distinguish a common fictional tale from reality, but AI can't. If a computer encounters a critical mass of references to a fictional story, it will mistake the story for reality.

Computers can't make a judgment call

Computers can't currently, nor will they in the near future, make a judgment call. So, for example, say that your personal assistant app is configured to tell you what diagnostics you should request during your next medical checkup, based on everything it knows about you: your sleep patterns, age, medical background, and lifestyle. As part of the decision-making process, the app also taps into a national disease database, and it finds in your area a high incidence of a particular type of cancer. So the app adds that screening to your list. You don't question why the app recommended these diagnostics, because this kind of routine checkup seems like something you can trust it to figure out.

What the app didn't tell you—indeed didn't know to tell you, because it wasn't programmed to tell you—was that for the past six years, people in your neighborhood have been diagnosed with that particular cancer, which is caused by waterborne contaminates.

Say that the app-recommended screening saved your life, because it led to an early cancer diagnosis. So the app solved a big problem for you. But it didn't solve the larger problem, because it didn't have the judgment to say, "By the way, you're drinking contaminated water. You should alert the authorities and, if you can, move to a different neighborhood." If it had, the municipality would have had the opportunity to fix the problem, and you could have considered moving. Instead, you, your family, and your neighbors continue to live in a dangerous situation.

AI decision-making is too complex to understand

Some systems make decisions for you in ways that you understand. For instance, you can assign many of the advertisements you see online to your Internet surfing habits. So if I look at a pair of red shoes on Zappos.com, the next time I'm on my news website, I'll see an ad for red shoes. Other systems make decisions in ways less transparent, but they're still relatively easy to understand. For instance, if you use Facebook regularly, you're allowing a computer algorithm to decide what (and whom) you see. Facebook shows in your feed only those posts that their algorithm deems interesting to you. On average, you see about a third of your friends' posts,[9] not because you didn't have time to look at them, but because Facebook decided you wouldn't want to see them. Although the inner workings of these algorithms might not be understood, the fact that these sites are using the data they collect about you to make choices on your behalf is pretty clear. Still other systems' decision-making processes are so complex that people can't begin to understand how they work.

Before you trust a recommendation an app makes, often you want more information. In the case of navigation apps, you can see a map with yellow and red roadways, indicating traffic you'd want to avoid. In the case of dating apps, you get not only a profile suggestion, but

also details explaining why the person could be a match. How could this "suggesting" function expand? Perhaps as you get more serious with the person, the app could tell you something about your communication with the person, for example, how often you use words of affection in your texts and whether the number is rising or falling. Perhaps an app could tell which of your habits have changed—either for better or worse—since you met this person. In Harari's hypothetical system, your app might indicate that it knows your bias for a better-looking partner, but statistics suggest looks will become less important to you in the long term.[10]

Ideally, AI will be able to provide similar explanations for its decision-making. If computers can provide some rationale for their recommendations, even if you don't agree with the rationale, it will help you determine if you should trust the recommendation. For example, a study showed that amongst Facebook users, the couples with the more long-lasting relationships were those who had numerous second-degree friends in common, rather than first-degree friends.[11] Researchers presume this indicated that spouses who have more second-degree friends in common have more common interests. But are these the real reasons behind this statistic? I don't know, but if a dating app that had access to Facebook profiles were to say, "We think this person is a match because you have many second-degree friends in common and statistically we've found that to have an X percent correlation to compatibility," you could decide for yourself if the recommendation makes sense to you. In order for people to feel comfortable with recommendations and maintain agency in a decision-making process, data scientists may have to create ways for computers to give a simplified explanation of the recommendation process, even when the calculations are so complex that they're not easily understood.

However, that's easier said than done. In many cases, that's just not how AI works. *Neural networks,* one form of machine learning, improve in the same way people do: by experience. For example, a neural network given images of skin diseases was able to more accurately diagnose future skin disease than experienced dermatologists.

Nobody gave the system rules for identifying melanomas. The data scientists just fed the system thousands upon thousands of images with the diagnosis for each image.

"The system isn't guided by an explicit store of medical knowledge and a list of diagnostic rules; it has effectively taught itself to differentiate moles from melanomas by making vast numbers of internal adjustments—something analogous to strengthening and weakening synaptic connections in the brain. Exactly how did it determine that a lesion was a melanoma? We can't know, and it can't tell us," writes Siddhartha Mukherjee[12], noting that this isn't any different from how a doctor's diagnosis might look. The doctor could give some specifics, but the actual answer is that the doctor has seen many, many cases, and she recognizes these symptoms as similar to those.

Neural networks are designed to improve themselves. Whenever a misdiagnosis happens, it can be fed back into the system, which then is able to make even better decisions in the future. Feasibly, these systems can find out they made a mistake automatically. With access to a patient's records, if the patient develops a disease three months or even ten years later, a machine-learning system can match that information to discover it made a misdiagnosis. Unlike a human doctor, the machine remembers each decision and isn't embarrassed to admit its mistakes.

The same kind of learning can be applied to many systems you use for your own decision-making. If an app told you that you would love some particular profession, and it turned out to be wrong, it could go back and analyze your data against other people who had been given similar misinformation and give better recommendations in the future. Still, it would be hard for this system to describe its decision-making in a simple way. Currently some startups are trying to address this challenge. Their attempts may make an app's recommendation seem more reliable, but as systems become more complex and the data involved becomes richer, in my opinion the best they'll be able to do is to explain a small piece of the multilayered rationale behind a recommendation.

AI presents a unique security risk

Delegation of decision-making into the digital space brings another aspect of scalable crime. When a hacker accesses your computerized agent, that one hack method could enable similar access to the agents of anyone who uses the same service. Therefore through a single, sophisticated hack, the attacker could gain control of many devices simultaneously, and each of those devices would have all of the details and access delegation for each person it serves.

Cybersecurity expert and former CTO of the Israeli Cyber Intelligence Unit 8200 Assaf Mischari describes how a phishing attack could look. Today, phishing emails are sent to individuals, and some people are careless enough to click on a link and proceed through the hacker's staged process into the trap. Such an attack is simple to create, yet it's success rate is limited. Think of a scenario where your computerized agent is independently receiving, responding, and initiating emails for you like renewing your insurance upon receiving a reminder email from your insurance company. A much more sophisticated phishing attack could target your agent and run it into a trap. That means everyone else with the same brand of automated assistant could be susceptible to the same email phishing attack. These security weaknesses must be addressed before anyone will entrust their credentials to an agent.

How Much Will AI Change Society?

Sociologically, data is becoming more and more important. So important in fact that thinkers in this field are calling it a new religion. Harari sees "data-ism" on the rise. Just as humanism replaced theism, Harari says, dethroning God as the ultimate decider of fate and placing human interest in that seat, data-ism could overtake humanism.

Harari explains, "Traditional religions assured you that every word and action was part of some great cosmic plan and that God watched you every minute and cared about all your thoughts and feelings." In

contrast, "Data religion now says that your every word and action is part of the great data flow, that the algorithms are constantly watching you and that they care about everything you do and feel."[13]

Data-ism believes that life's decisions should be made by intelligent computing systems, based on facts. Harari points out that, even if you're a humanist, research has shown humans aren't particularly good at recognizing or understanding their own feelings. Using the example of an election, he suggests a computer could calculate my reactions and emotions to a particular party's performance over the entire time that it's in office, based on data like my posts on social media or my physical state when I read a particular article. Then, when the elections come around, instead of being biased by what I have felt in the last month, an AI system could cast a vote that better represents my opinion. Given that campaigns often make dramatic moves right before an election in order to sway votes, Harari's hypothesis makes sense. Perhaps it feels like you would be losing your free will if you didn't cast your own vote, but if a data-driven vote better represents your actual desire, maybe letting your agent vote would give you more, not less free will?

Whether you buy into data-ism or not, the evidence is clear: computers will continue to change how people make decisions, and in that way they will continue to change who people are as individuals. How will these developments impact culture? Already, nationally and internationally, it's evident that there's an enormous divide between the affluent and those who live below the poverty line. Access to technology helps some people continue to thrive while lack of access bars too many people from job markets, educational opportunities, and other essential resources. As developments in AI and data-driven decision-making enable those people with access to enhance their moods, thinking capacities, physical capacities, and performance in school and at work, those without access will be left further behind.

Today, only a small part of humanity can afford DNA testing and the cost of a preventive operation like Angelina Jolie had. Few people can afford the kinds of equipment top athletes get to use. As these technologies become more powerful and expensive, the gap between the haves and the have-nots will continue to widen.

Who Is Behind the Computers?

As you hand decision-making tasks over to computers and as those computers become parts of your identity, you may wonder which interests beyond your own are influencing the decision-making process.

Computer services are offered by organizations that have their own goals. This brings new questions that need to be considered.

What biases will be built into AI technologies by the businesses that create them? Even if the interest of the business is to serve me best, what are the set of values that guide them? Are they specific to a certain culture? Is it always good or bad for business when people feel happier? What methodology of happiness creation is better for business? How will the values built into these systems influence people's identities?

Small biases embedded in algorithms can, over many iterations, result in, well, big biases. It's not a major issue if Alexa is biased to sell you one brand of plumbing fixture over another, if both cost the same and perform equally. Over time, though, could such gaming of the system influence which businesses succeed and which fail? Could seemingly subtle biases lead to major changes in how people lead their lives?

Whether it's an enhanced human, an AI computer, a government, or a business, the underlying question remains the same: will you trust them?

You Will Delegate

Initially, when GPS systems became available, they gave route options, and they recommended the shortest route. Users were able to consider the options and choose their preferred route, based on, say, their knowledge of traffic patterns or a road closure. Today, navigation systems, like Waze, are tracking traffic and recommending the fastest route not only based on current data but also using predictive algorithms. If you've ever overridden today's navigation systems and

chosen a different route (and most people have, at least once), odds are you've regretted it. You no longer even think about exercising your own opinion when it comes to navigation decisions because, face it, Waze is always right.

Right now, reading this chapter, it might seem implausible that you would give an automated system control over significant decisions in your life. But we as society have seen over and over again that humans have a gap between how they think they will act in the future and what actually happens. At one time, it seemed like people would be careful about sharing their data, but today, people readily share their data both privately and publicly, often with little consideration. I'm not so sure a teenager will pause to contemplate philosophical considerations before using an app that could improve his chances of choosing a trend-forward outfit for his next night out. Adults also will be exposed to temptations. People are constantly facing decision-making challenges, from a first-time parent of a five-year-old experiencing social challenges, to a homeowner choosing the right tree to plant in the backyard. So, when wondering what decisions you might delegate to a computer in the future, "a lot" might not be such a far-fetched answer. Today, people allow an app to decide the best driving route or the best reading material. Tomorrow, will people allow an app to tell them what job is suitable, where they should live, whom they should marry, or how many children to have?

This future may sound ominous, but the new reality won't fall upon us as a surprise. On the contrary, we will consciously choose it and proactively authorize computers to manage parts of our identities, when doing so serves us.

Endnotes

1. https://tim.blog/2016/01/30/naval-ravikant-on-happiness-hacks/
2. Steve Lohr, *Data-ism: The Revolution Transforming Decision Making, Consumer Behavior, and Almost Everything Else.* HarperCollins, 2015.
3. https://www.nytimes.com/2016/05/29/opinion/sunday/why-you-will-marry-the-wrong-person.html
4. https://www.ft.com/content/50bb4830-6a4c-11e6-ae5b-a7cc5dd5a28c?mhq5j=e1
5. Yuval Noah Harari. *Homo Deus: A Brief History of Tomorrow.* HarperCollins, 2017.

6. https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-2.html
7. http://www.newyorker.com/magazine/2017/04/03/ai-versus-md
8. https://www.newscientist.com/article/dn20128-better-than-human-whats-next-for-jeopardy-computer/
9. http://www.businessinsider.com/35-percent-of-friends-see-your-facebook-posts-2013-8
10. https://www.ft.com/content/50bb4830-6a4c-11e6-ae5b-a7cc5dd5a28c?mhq5j=e1
11. Lohr. *Data-ism.*
12. http://www.newyorker.com/magazine/2017/04/03/ai-versus-md
13. Harari, *Homo Deus*.

Epilogue

"We wanted flying cars, instead we got 140 characters," Peter Thiel famously said.[1] This complaint about the state of technology touches on two human tendencies: 1) We as human beings love trying to predict what the future will look like; 2) We aren't very good at it. Often our predictions begin with the kinds of technology we think we might see. Much has been said about the pace of technological progression—changes that may have taken a hundred years once upon a time, now occur in a decade or less. We must take into account that acceleration when trying to gauge the future. But it's also important to consider the changes that *follow* new technology. Often the developments that surprise us aren't technological, but rather behavioral. Who would have guessed that we'd be so excited about obsessively announcing to the world everything we do and think? Yet multiple companies gained billions of dollars in value, just by enabling people to share content with each other.

The innovation cycle begins with technology and continues with business initiatives, which bring new experiences. As these experiences become part of our daily lives, our behavior changes, and so do we.

Before mass production, we as a society couldn't have imagined how much we would want to acquire and consume. Before smart mobile devices, we didn't know we would develop a need to be constantly entertained. Before identity digitalization, we didn't think corporations would, in many ways, know us better than we know ourselves. Now, with the rise of data-driven decision-making and artificial intelligence (AI), we must ask ourselves, will we lose our independent thinking?

How Will the Digitalization of Identity Change Us?

The choices we make form a big part of our identities. Our destinies are determined largely by our decisions. So what happens when we start making choices based not on our individual perspectives, but

based on data? If we let algorithms make our decisions, will we sacrifice our individuality? Instead, should we continue to make subjective and emotional choices to maintain our humanity?

We are entering into the future of identity. Technology enables the creation of a digital doppelgänger for each of us, translating who we are and what we do into data. Businesses find ways to create value out of that, for them and for us. As consumers, we take advantage of this behavioral data in multiple ways, including using it to inform our decisions, as we see in the Quantified Self movement. This movement, in some shape or form, will become ubiquitous. We'll turn on systems that measure many aspects of our daily lives in order to improve them, and doing so will transform us.

The data will tell us who we are and show us the results of our choices, unfiltered. Our behavior patterns and triggers will become visible, opening opportunities to resist natural biases. In this way, we can become much more analytical, rational, and controlled in our decision-making.

We are marching toward an era of extreme analytic thinking. For those of us who choose to practice it, data-based decision-making can lead to a more structured, predictable, and sensible world. Analytic thinking can help us feel more confident in ourselves—we'll harbor less regret when we know we have taken the rational path. In fact, eighteenth-century philosopher Immanuel Kant would say reason is the key to moral behavior. (Of course, he knew nothing of big data.)

Theoretically, we as consumers will be able to use analytic thinking to help understand which decisions bring the greatest happiness, thereby enabling us to create more satisfying lives. How many times have you done something and said to yourself, "I forgot how much I enjoyed (or hated) this activity"? In an analytic world, you'll be able to bypass these flaws in your memories instead making decisions based on data about your past experiences and desired emotions.

In the near future, AI will connect with our digital identities and will further influence our decisions, potentially helping us avoid bad choices, as parents direct their children. With this assistance, those of us who are so inclined could enhance our individual lives, while also making more ethical and environmentally conscious decisions.

Figure A-1: *Does this look familiar?*

As decision-making products come to know us, they'll be able to advise us to make not only the most efficient choices, but also the choices that best match our values. Already we as consumers are seeing navigation systems that give us options to take the most scenic route, the safest path, or the most ecofriendly transportation to our destinations.

But There Is a Problem Here

The digitization process is unstoppable and, I contend, for the good. However, there is a problem. The more optimized our decisions are, the more similar we will become. The data shows us only a superficial

reflection of ourselves. So, to maintain our individuality, at the same time analytic capabilities turn us into more computer-like beings, we must also develop our minds organically. How? Some of us develop by exploring our souls in therapy; others meditate, observe nature, create art, or play music.

No matter how we choose to foster it, our organic development will be critical to keeping abilities we might lose without noticing. For example, dating apps may help us find a partner, but still we must learn how to create healthy relationships with our partners.

By continuing to grow as human beings, including spiritual and emotional sides, we'll take responsibility for aspects of our individuality that can never be replaced by a machine. No matter what technologies develop, we can always explore, make new choices, and grow. Most importantly, when we stay in touch with our intuitive side, we'll know when to make our own decisions, even when the data suggests differently.

Even when AI solutions learn to make decisions based on our values, they can't replace our individuality. Because beneath our values lie passions, fears, expectations, and hopes—the truly unique aspects of our personalities. When we know what drives us, and we remain dedicated to our organic development, we can use the tools of the digital future without worrying about computers taking over our lives.

We often have heard that the journey is more important than the destination. As we move deeper into the era of digital identity, now more than ever, the engine that drives us will be far more important than both.

Endnote

1. http://som.yale.edu/blog/peter-thiel-at-yale-we-wanted-flying-cars-instead-we-got-140-characters

WE HAVE ARRIVED AT A CRISIS POINT IN THE DIGITAL IDENTITY REVOLUTION, WITH FAR-REACHING IMPLICATIONS FOR INDIVIDUALS AND BUSINESSES. TRUST IS BECOMING THE NEW CURRENCY.

Companies are collecting, storing, and sharing ever more personal information about customers. Big data is changing the rules of the marketing game. Government privacy regulations are attempting to assert some control. Major data breaches have people concerned about the security of their data, while at the same time individuals share increasingly more private information with apps that promise to enhance their health, finances, job performance, and social lives. What is happening here? Have we unleashed chaos or unprecedented possibilities? Maybe both. *The Digital Identity Crisis* asks some serious questions about the risks and opportunities we face, now and in the near future.

- **What does privacy mean in the digital age?**
- **Who owns digital identities?**
- **How can rich customer data disrupt the way businesses operate?**
- **How has trust become the most essential factor in relationships between companies and their customers?**
- **How can businesses build trust-based relationships with their customers?**
- **Why are cybercriminals interested in personal data?**
- **Will passwords ever die?**
- **Can digital identity improve the quality of life? At what cost?**
- **How will AI influence identity?**

ROOLY ELIEZEROV

is president and cofounder of Gigya, the industry leader in customer identity management solutions, recently acquired by SAP. A veteran entrepreneur since the Internet's earliest days, Eliezerov has explored identity through the lenses of art, architecture, sociology, and of course, technology. From his vantage point at Gigya, over the past decade he has seen broad-reaching changes in the ways people share their personal data, the way businesses engage their customers online, how governments regulate identity data, and the extent to which data-based technologies transform people's lives.

BUSINESS & ECONOMICS / Industries / Computers & Information Technology

S19.95 USA

ISBN 978-1-119-47985-7

51995

9 781119 479857

Cover Design: Wiley

Cover Image © Getty Images

Author Photo: Courtesy of Gigya

WILEY

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.