TESTIMONY OF DR. ALEKSANDR KOGAN
BEFORE THE
COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION
SUBCOMMITTEE ON CONSUMER PROTECTION, PRODUCT SAFETY, INSURANCE, AND DATA SECURITY
UNITED STATES SENATE
JUNE 19, 2018

Chair and Members of the Subcommittee:

The past several months have seen a myriad of allegations and claims made about myself, Cambridge Analytica, and Facebook in relation to a project carried out in the summer of 2014. Some of this information has been accurate, while other aspects have been speculation, exaggeration, or misinformation. My hope is to help the Subcommittee understand what is and is not legitimate from the various narratives that have emerged. I also wish to give my current thinking as to the important broader issues raised by the controversy surrounding Facebook and the privacy of user information.

My Background

I am a social psychologist whose academic work focuses on well-being, kindness, and compassion. To study these topics, my lab and I have used a variety of methods, including surveys, behavioral studies, and social media. I received my B.A. degree in Psychology from the University of California, Berkeley, in 2008 and a Ph.D. in Psychology from the University of Hong Kong in 2011. Since 2012, I have been a Research Associate and University Lecturer at the University of Cambridge (the "University") in the Department of Psychology. At the University, I have conducted research, taught classes, and supervised graduate and undergraduate research work through the Cambridge Prosociality and Well-being Laboratory (the "CPW Lab")—which I founded. All of my academic work was reviewed and approved by the University's ethics committees.

My Collaboration with Facebook and the CPW Lab App

In early 2013, I began collaborating with Facebook on studies aimed to understand how people around the world connect and express emotions. Throughout 2013, Facebook provided me with several macro-level datasets on friendship connections and emoticon usage. These were aggregated datasets, typically at the country level (e.g., number of friendship connections between USA and UK), which did not contain specific information about individual Facebook users. Using this data, members of my lab began writing papers together with Facebook personnel.

During this active collaboration with Facebook, I created a Facebook app, which I called the CPW Lab App (after the name of my lab), in order for us to collect individual Facebook users' data to pair with the aggregated data Facebook had provided directly. For studies based on data derived from the app, we asked participants to complete a survey and provide information from their Facebook accounts by logging in through the Facebook application portal. The terms of

service of the CPW Lab App were contained in a link on the Facebook application portal's login page. The terms of service indicated that the data would be used for academic purposes. Data derived from the CPW Lab App was not provided to SCL.

The GSR App and Data Collection

I was introduced to SCL through a Ph.D. student at the University in winter of 2014. He introduced me to Chris Wylie, who represented SCL at the time. Our conversations began with Mr. Wylie detailing his experiences working for the Obama campaign. He asked me to provide survey-consulting services to SCL, including collecting data from Facebook and generating personality profiles.

To do the project, a fellow University research psychologist and I registered a company, Global Science Research ("GSR"), in the UK. Mr. Wylie held himself out to us as a data law expert, having studied law at a London university, and he served as our guide on how to be compliant with legal requirements and prohibitions.

Before we started collecting data from survey participants, GSR changed the name and terms of service of the CPW Lab App. The terms of service were provided to us by Mr. Wylie. References to academic use and the University were deleted from the terms of service, and the name of the application was changed from "CPW Lab App" to "GSR App." GSR also changed the terms of service of the application to reflect the expected use of the data. When individuals who participated in the survey logged into Facebook through the GSR App portal, Facebook presented a link to the GSR App's terms of service, which informed each participant as follows:

> [i]f you click "OKAY" or otherwise use the Application or accept payment, you permit GSR to edit, copy, disseminate, publish, transfer, append or merge with other databases, sell, licence (by whatever means and on whatever terms) and archive your contribution and data . . . and grant GSR an irrevocable, sublicenceable, assignable, non-exclusive, transferrable and worldwide license to use your data and contribution for any purpose.

The terms of service also informed each survey participant that GSR would collect "any information that [the participant] choose[s] to share with us by using the Application. This may include, inter alia, the name, demographics, status updates and Facebook likes of your profile and of your network."

After the participants entered their Facebook credentials into the GSR App Facebook login portal, they were taken back to the third-party survey vendor's website to complete the survey. GSR collected data from the survey participants and their friends whose Facebook privacy settings were set to allow the participants access to their information. The data collected from participants and friends included, if available, an individual's name, birth date, location (city and state), gender and the Facebook pages each user had "liked." As with the CPW Lab iteration of the application, information was collected from friends whose Facebook privacy settings were set to provide the survey participants access to the friends' "likes" and demographic information.

In the end, approximately 30 million personality profiles based on this information, plus a limited amount of demographic data and certain "likes," were transferred to SCL.

In the latter part of 2014, after the GSR App data collection was complete, GSR revised the application to become an interactive personality "quiz" called "thisisyourdigitallife." The commercial portions of the terms of service that had been added to the GSR App were not changed. The thisisyourdigitallife App was used by only a few hundred individuals and, like the two prior iterations of the application, collected demographic information and data about "likes" for survey participants and their friends whose Facebook privacy settings gave participants access to "likes" and demographic information. Data collected by the thisisyourdigitallife App was not transferred to SCL.

Micro-targeting on Facebook

A point of confusion has been whether the data we collected would be useful for micro-targeting ads on Facebook. I believe the project we did had little to no use for someone wanting to run targeted ads on Facebook. The Facebook ads platform already provided SCL with many tools to run targeted ads with little need for our work—in fact, to this day, the platform's tools provide companies a far more effective pathway to target people based on their personalities than using personality profiles for Facebook users developed by myself and my fellow social psychologists.

Negative Public Reaction

In reflecting on the SCL project and the public reaction from media reports, I perceive two primary reasons for public concern. First, people may feel angry and violated to the extent that their data may have been used as part of a mind-control effort; that somehow SCL had figured out their inner demons, weaponized the internet, and used this ability to dupe them into voting a particular way when otherwise they would not have. This concern rests on an incorrect premise about the data and its utility. I believe there is almost no chance this data could have been helpful to a political campaign—and I still have not seen any evidence to indicate that the Trump campaign used this dataset to micro-target voters. The arguments for its utility fall apart under legitimate scientific scrutiny, and I would be happy to talk in greater detail about the underlying issues.

Second, regardless of the data's efficacy, people may feel angry and violated to the extent that their data could have been and was accessed by others without their appreciating the actual access they were giving third parties. This is an understandable emotional reaction upon realizing how much data was being conveyed directly and indirectly. I'm very regretful that I did not better anticipate this reaction. If the heart of the controversy lies in this second issue, I believe it points to a much broader problem with how companies interact with consumers in the tech space—in particular, the conduct of companies whose business model is predicated on digital advertising.

Roots of the Controversy

Given what we now know, I believe that a situation like the present one was inevitable. For decades, a shift occurred in how consumers interacted with companies. The interaction used to be quite simple: Company gives us product, we give company money. It was typically impersonal and arms-length. Then digital marketing came into existence, and new companies arose with a new formula: They give us technological products and services, not in exchange for money, but in exchange for intimate details about ourselves that we are willing to share. It is our photos, thoughts, emotions, and connections. These things become valuable to companies, and they use this information to monetize us to their actual clients: Advertisers. We became the product. This new relationship between company and consumer is extremely complicated and personal. But the primary vehicle we have had to manage this relationship has been a "terms of service" document that is often unhelpful to the average consumer.

A core aspect of many of these documents is the idea of blanket consent: The consumer gives broad rights for a company to do whatever it likes with the data. For instance, here is an excerpt from the Facebook Terms of Service from 2014 (section 2.1):

> For content that is covered by intellectual property rights, like photos and videos (IP content), you specifically give us the following permission, subject to your privacy and application settings: you grant us a non-exclusive, transferable, sub-licensable, royalty-free, worldwide license to use any IP content that you post on or in connection with Facebook (IP License).

Blanket consent such as this is standard industry practice—Facebook is hardly alone in using language like this. It was and still is normal for companies to ask consumers to grant them broad rights to do whatever the companies like with the data. The widespread use of such agreements, in my view, created a near certainty that at some point consumers would be unpleasantly surprised by the specific way a company used their data. Consumers may have technically agreed to broad usage, but nonetheless the feelings of outrage and being wronged that have been released by the Cambridge Analytica controversy are the inevitable response.

Tackling the Issue

Trying to fix this problem will not be simple. My expertise provides one perspective on the issues—and so I very much look forward to a dialogue with others to contribute to public awareness and broaden my own understanding of the various forces at play.

As a general matter, businesses are typically able to respond to consumer needs and dissatisfaction, but this may be difficult for those businesses that have a strong financial incentive to collect more and more data. Companies whose main revenue comes from ads are typically selling advertisers on the idea that the companies can find the right person, in the right place, at the right time, and serve them the right ad. This may act as a barrier to change.

In fact, digital ads, in particular on social media, are far less effective than one might think. For example, in the only paper I know of to have tested the idea of tailoring Facebook ads based on

people's personalities, the researchers found that less than 1 in 100 actually clicked the ad. And of those few people, fewer than five in 100 did what the ad wanted them to do—buy a product. Facebook and companies like it are under enormous pressure to do better. That often means they want more and more data about us to build better and better models. Enhancing consumer consent, such as requiring opt-in consent, jeopardizes their ability to collect more and more data—in fact, opt-in consent will likely strip away much of the data the companies have today. So I believe companies are likely to strongly resist making changes themselves for fear that others will not follow suit and they will be at a competitive disadvantage.

A second important issue is the role of autonomy and individual responsibility. We as Americans strongly cherish our ability to choose—we see it in our system of government and in our approaches to consumer goods. So I believe any solution needs to be respectful of people's right to choose. Critically, however, this ability to choose must be predicated on being informed about (a) what a person is consenting to and (b) the risks of giving consent. In short, people must give true informed consent. However, the kinds of blanket approaches to consent used by major data-monetizing companies like Facebook run counter to true informed consent.

Some Possible Responses

In my view, thinking through the means of achieving informed consent is the key to avoiding a future Cambridge Analytica situation. I see a few ways we can start going down this road. First, I would suggest we find ways to give consumers the information they need about how their data will be used—and these planned uses need to be specific rather than abstract and general. Rather than a broad license, companies could outline specifically how the data will be used (e.g., for advertising about consumer products, advertising about political campaigns, building models of your values and preferences). This will avoid consumers being caught off guard later on.

Second, consumers may need some kind of risk assessment of what could go wrong—companies should be upfront about what dangers exist in placing data on their platform and what they are doing to protect against those risks.

Third, we should consider requiring that consumers give opt-in consent. It should not be taken as a given that a person agrees with everything unless they say they disagree; it is important for people to make an affirmative act of agreement to demonstrate the making of a decision.

Fourth, we could consider requiring consent to be given to each specific point in the consent document rather than globally. This would give consumers the ability to choose what they are and are not comfortable with, and also incentivize companies to provide shorter, more readable documents.

Fifth, we should look for ways to prevent or at least discourage consent being given without consumers reading—or even seeing—the disclosures. Facebook currently allows consumers and developers alike to sign up to their main service and their apps without being required to see the terms of service. They are hard to find, easy to miss, and do not always require an affirmative manifestation of consent. This, in my view, runs counter to the notion of true informed consent from consumers.

Conclusion

In sum, my view is that we should think hard about finding ways to empower consumers, giving them the ability to make more informed decisions about how their data is used. This, I fear, cannot be left entirely to companies and consumers to work out among themselves as business interests may run counter to consumer privacy interests because of present revenue models.

Sincerely,

Aleksandr Kogan