

Written evidence submitted by Aleksandr Kogan

Over the past several months, your committee and the general public have seen a number of allegations and claims about myself, Cambridge Analytica, and Facebook—some of these claims have been true, while many others have been speculation, exaggeration, or misinformation. I am very much looking forward to the opportunity to help set the record straight. With that goal, I write to you today to provide an honest account of events to the best of my understanding, knowledge, and memory in preparation for my testimony on April 24th, 2018.

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 use 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 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 and direct.

With relevance to social media, the CPW Lab’s research used information gathered from individuals through multi-question online surveys. The participants in these studies were recruited through third-party online survey vendors, such as Amazon Mechanical Turk and Qualtrics, and are individuals who have agreed through those vendors to take online surveys for money. Participants are paid a small sum (usually a few dollars) in exchange for completing each survey. All of my academic work was reviewed and approved by the University’s ethics committees.

The Facebook App and My Collaboration with Facebook

In early 2013, I began collaborating directly with Facebook on studies aiming to understand how people connect and express emotions around the world. Throughout 2013, Facebook provided me with several macro-level datasets on friendship connections and emoticon usage. These datasets were aggregated, typically at the country level (e.g., number of friendship connections between USA and UK).

As my lab began writing papers with Facebook using this data, 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 level data to pair with the data Facebook had provided directly. In studies where we used the app, we would ask participants to complete a survey and to provide us with information from their Facebook accounts by logging in through the Facebook application portal. The terms and conditions of the CPW Lab App were contained in a link on the Facebook application portal’s login page. The terms and conditions indicated that the data would be used for academic purposes.

Approximately 5,000 to 15,000 survey participants logged into Facebook through the CPW Lab App, and their data was collected in an anonymized manner. The CPW Lab App also collected information that members of the participants’ Facebook friend networks (“friends”) had shared with the participants, including demographic information and the Facebook “pages” that the friends had publicly “liked.”

Friend information was collected for individuals whose Facebook privacy settings gave the survey participants access to the friends' "likes" and demographic information. For a few studies, we also collected private message data (I estimate approximately 1000-2000 users participated in these studies). Private message data was collected only from participants—not friends—and we made it explicitly clear in the consent form at the beginning of the study we would be collecting this information. The private messages were of interest in our studies of how people expressed emotions online. Data collected through the CPW Lab App was housed at the University of Cambridge and used for academic purposes; it was not provided to SCL Group. As with all of my research, the research that used the app was first ethics approved by the University.

SCL Relationship

I was first introduced to SCL through a PhD student at U of Cambridge in February 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, and the desire to share commercial datasets with academics. He asked me to potentially provide survey-consulting services to SCL.

Eventually the conversation turned towards personality and Facebook, and at that point I introduced Mr Wylie to David Stillwell, who had collected a large Facebook personality dataset called myPersonality. Mr Wylie was interested in acquiring the myPersonality dataset from Dr Stillwell, but after considering the proposition for a few months, Dr Stillwell decided against it on the grounds that he had collected the data for academic purposes—and that is what he warranted to his users. At this time, Dr Stillwell, Dr Michal Kosinski (a close collaborator of Dr Stillwell's), and I proposed making personality predictions for SCL. Dr Stillwell and Kosinski had already developed a model to make predictions from likes, and I had experience using the Facebook login app from my previous. This idea of predicting personality from page likes became the foundation of the project that I did with SCL. Eventually, Dr. Stillwell and Kosinski were removed from the project because they requested \$500,000 to work on the models—a request SCL ultimately felt was unreasonable given the unproven commercial nature of the models. And so I was asked to handle both the data collection and modeling.

To do the project, a fellow University of Cambridge research psychologist and I registered a company, Global Science Research ("GSR"). In June 2014, GSR entered into a data and technology subscription agreement with SCL Elections Limited ("SCL") (the "Agreement"). GSR agreed to provide SCL with specified demographic and personality data for individuals in 11 US states. In return, SCL agreed to pay the cost of collecting the data, which consisted chiefly of fees paid to survey participants (about three to four dollars per participant). SCL's payments under the Agreement, made directly to the Qualtrics, amounted to about \$800,000. During the Summer of 2014, Qualtrics recruited between 200,000 and 300,000 participants to take a survey developed for the purpose of collecting particular data.

Throughout the project, Mr Wylie worked as our guide on how to be compliant with all legal frameworks and requirements. Given that we were academics with little money, were not going to be getting paid from the Agreement, and were assured that Mr Wylie was a data law expert given his previous experience, we relied on Mr Wylie exclusively for guidance on compliance.

Before we started collecting data from survey participants, GSR changed the name and terms and conditions of the application. The terms of service were provided to us via email in early June by Mr. Wylie. References to academic use and the University of Cambridge were deleted from the terms and conditions, and the name of the application was changed from "CPW Lab App" to "GSR App." GSR also

changed the terms and conditions 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 and conditions, 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 automatically 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 late 2014, GSR provided SCL with the data and analyses called for by the Agreement. This consisted of (i) demographic information for survey participants and their friends in the 11 specified states; and (ii) personality scores and a limited number of predictions for survey participants and friends based on the collected data.

SCL later requested that we provide data and analysis for survey participants and their friends for all 50 states. In early 2015, we provided the additional data requested by SCL. This second set of data contained the same types of demographic information, survey responses, but fewer personality analyses than those that had been provided for the original 11 states that had been turned over pursuant to the Agreement. In addition, for the second set of data, SCL requested information on whether the friends and survey participants had "likes" for about 500 pages specified by SCL, which included some political figures and celebrities. GSR provided this additional "like" data, which amounted to about four percent of each friend's "likes." SCL also requested and received information on who people were friends with (social network connections). In return for the second set of data, SCL paid GSR £230,000. These funds were used by GSR for research, development, administrative costs, and professional services. I did not receive any salary from GSR at any point during its operation.

Other entities we shared data with

GSR also transferred an anonymized copy of the GSR App data to the CPW Lab (for research purposes), and some derivative information (for example, psychological trait predictions) from some of the GSR App data to academic research colleagues at University of Toronto.

Lastly, GSR entered into an agreement with Eunoia (Chris Wylie's company) in the summer of 2014 to provide GSR App data to Eunoia in return for getting other commercial datasets from Eunoia. Under this

agreement, GSR provided Eunoia with a copy of all of the GSR App data for people who reported their location in the United States along with GSR personality analyses on some of the data. After Eunoia failed to provide the promised data to GSR, GSR instructed Eunoia to destroy the GSR App data that had been transferred.

For clarity, there is a substantial difference between the data SCL and Mr Wylie's company were provided. SCL was never given, at least by GSR, access to the raw Facebook data containing all of the Likes. SCL received only demographic information (if available, name, birth date, location (city and state), gender) and personality predictions and, later in 2015, the limited set of 500 page likes specified in 2015, representing 4% of the overall Likes. This is in contrast with the contract with Mr. Wylie's entity Eunoia, where Eunoia received all of the page like data as well as dyads.

Thisisyourdigitallife

In the latter part of 2014, after the GSR App data collection was complete, we at GSR revised the application to become an interactive personality "quiz." For this purpose, the app was renamed "thisisyourdigitallife." The commercial portions of the terms and conditions quoted above were not changed. Participants could authorize the application to view their profiles, and the application would analyze the participants' "likes" and provide the participants with insights regarding their personalities. Participants were not paid to use this iteration of the application.

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 provided to SCL.

Accuracy of personality scores

One of the biggest points of confusion has been how accurate the personality scores we provided to SCL were. The truth is that the scores were highly inaccurate. We estimate that we were right about all five traits for about 1% of the people—in contrast, we were wrong about all 5 traits for about 6% of the people. Looking at accuracy another way, we found that the scores were more accurate than a random guess, but less accurate than assuming everyone is average on every trait. I'd be delighted to discuss this in more depth when I testify with a slide deck to illustrate this point.

Need for the dataset for micro-targeting on Facebook

A second point of confusion is whether the data we collected would be useful for micro-targeting ads on Facebook. I believe the project we did makes little to no sense if the goal is to run targeted ads on Facebook. The Facebook ads platform provides tools and capability to run targeted ads with little need

for our work—in fact, the platform's tools provide companies a far more effective pathway to target people based on their personalities than using scores from users from our work. I again look forward to discussing this in greater detail in front of the committee.

Aleksandr Kogan
April 16, 2018