

# Digital, Culture, Media and Sport Committee

## Oral evidence: Fake News, HC 363

Tuesday 24 Apr 2018

Ordered by the House of Commons to be published on 24 Apr 2018.

Watch the meeting

Members present: Damian Collins (Chair); Julie Elliott; Paul Farrelly; Simon Hart; Ian C. Lucas; Christian Matheson; Brendan O'Hara; Jo Stevens.

Questions 1770-2086

### Witness

I: Dr Aleksandr Kogan, Senior Research Associate, Department of Psychology, Cambridge University.

Written evidence from witness:

– [Dr Aleksandr Kogan](#)

### Examination of witness

Witness: Dr Aleksandr Kogan.

Q1770 **Chair:** Welcome to this further evidence session of the Digital, Culture, Media and Sport Committee, as part of our inquiry into disinformation and fake news. We are delighted to welcome Dr Aleksandr Kogan to give evidence to the Committee this morning. Although there is a wide range of topics that I am sure will come up, and this is a complex matter, I think there are two or three quite clear areas of interest to the Committee, which I am sure will come out during questioning. First, there is what has been called the Facebook data breach, in which you, Dr Kogan, have been seen as a central player. We want to understand more about that, and in particular Facebook's knowledge of the work that you were doing, and the extent of that work. Secondly, we want to understand more about the use that Cambridge Analytica made of the work that you did and your relationship with them. Thirdly, we want to



## HOUSE OF COMMONS

understand a little more about your work in general, your academic research, and the work you have done for other people.

Looking at that first area—trying to understand a bit more about the way in which you worked with Facebook data, and Facebook’s knowledge of the work that you were doing—I was very interested to read the written statement that you supplied to the Committee, which the Committee has now published and is available online through the Committee’s website for anyone to read. You said: “Throughout 2013, Facebook provided me with several macro-level datasets on friendship connections and emoticon usage.” Could you explain a little more about that? This was in the period before you developed the app that is at the centre of the Facebook data breach. What sort of information did Facebook provide you with?

**Aleksandr Kogan:** We were collaborating with a team at Facebook—I believe they worked on protect and care. The dataset we received initially was about every friendship made in the world between every country over a number of years, and it was broken down by month. It was how many friendships were created between the United States and the UK in, say, March 2011, how many friendships were made between the UK and Canada in April 2012, and things like that. When we say “macro”, we mean at the country level or in the aggregate.

Later on, they provided me with datasets about emotional expression. These are the emojis that we often see—the smiley face, the sad face, and things like that. This was also in the aggregate. It was a bit more segmented, by which I mean it was broken down by gender and age groups—how many smiley faces were used by, say, men who were 20 to 30 years old, and things like that. That was for a particular week in, I believe, August 2013.

Q1771 **Chair:** From the information that was given to you by Facebook, was it possible to break that down in any way? Could you extrapolate information about individuals?

**Aleksandr Kogan:** No.

Q1772 **Chair:** It was purely macro-level data.

**Aleksandr Kogan:** Exactly.

Q1773 **Chair:** Was that data useful to you in the work that you were doing? Did it help you to design other tools? Did it help you to design apps that you went on to work on?

**Aleksandr Kogan:** It was useful from an academic perspective. We were using it purely for the lab, and we were studying things like whether we can predict how much money is donated during natural disasters by different countries based on the number of friendships that exist between the two nations. It was really this macro perspective.

Q1774 **Chair:** How did you become introduced to Facebook to the extent that they were prepared to give you data like this to work with?

**Aleksandr Kogan:** My mentor and adviser from my undergrad days at UC Berkeley was consulting for them. He was working with a protect and care team to solve problems of people posting embarrassing photos and not taking them down. He was helping them to understand how we can better tackle that. The focus was on emotions and speaking to people on an emotional level: “Hey, this isn’t making me feel good. Would you mind taking it down?” My understanding was that this was really helpful to Facebook in combating that issue.

**Q1775 Chair:** Who was the person or people at Facebook who you worked with at this time on this project?

**Aleksandr Kogan:** There were a couple of user experience researchers who were on the protect and care team.

**Q1776 Chair:** Are you able to say who they were?

**Aleksandr Kogan:** I can, but I would prefer to do so privately—they are not really central to the story and I want to protect their confidence.

**Q1777 Chair:** That is understood and perhaps we can follow up on that separately. When Facebook gave evidence to us in February, it said that it did not give out any Facebook user data, but clearly it did, even if it was in macro form.

**Aleksandr Kogan:** I think that is tricky. Most data that we will look at around the world, and public data stats, comes ultimately from the individual. Even World Bank data or World Health Organisation data—ultimately, any economic activity is driven, at least at the start, by people. When you deliver to somebody you are going to get it at country level or organisation level, and I think Facebook is thinking about it in the same way. We are giving you aggregated information that was generated from users—once you go up to that level it is very difficult to go back down.

**Q1778 Chair:** I suppose that if you are giving people data on a macro level, the purpose of doing that is to help you design tools to target Facebook users on an individual basis who understand how you might do that.

**Aleksandr Kogan:** Not in this case. This was really an act of collaboration. The people I was working with had PhDs, so they had a strong interest in research generally. They were now working at Facebook as user experience researchers, but they maintained that academic passion. This was a chance for them to go and explore that academic passion. That was really the focus; there is really no conversation at this point about any Facebook tools or anything like that. It was really one way for us to benefit and build research.

**Q1779 Chair:** All research has a purpose, and as you said earlier, if you were looking at how you could encourage people from one country to give money to support disaster relief in another country, the methodology for doing that could be derived from the work you were doing looking at the interactions between Facebook users in different countries.



## HOUSE OF COMMONS

**Aleksandr Kogan:** I think that is quite downstream. We are doing very early basic science. Before we get to application there are a lot of steps there. In our case we are looking just at associations. There is really interesting potential for impact, but that is quite downstream and requires quite a bit of extra research. I do not know if Facebook would really be the vehicle for achieving this. I think Governments would be.

Q1780 **Chair:** But for downstream to exist there has to be an upstream. It has to start somewhere, and this would seem to be an area where it would start. From what you said now and in your statement, Facebook has not been clear when it discussed the way that the data it holds is used by outside organisations. It sought to draw a very clear distinction in its evidence to the Committee between Facebook user data that is accessed by developers through tools, and apps that they create and interact with, and data that Facebook holds. It said to the Committee that it does not share data that it holds on users—the only user data that is gathered by developers is done through tools that they have created. From what you said, I think it is not as clear as that.

**Aleksandr Kogan:** Yes. If your basis is that aggregated data is derived from users, and that in your mind would still count as user data, then I would agree with that statement.

Q1781 **Chair:** And Facebook clearly had a programme of working with academics and outside institutions, in many ways to try to help it understand the aggregated data based on its users.

**Aleksandr Kogan:** Yes, I would not go as far as ascribing that much intent, as far as its desire, but it is well documented that Facebook collaborates with researchers. You can just go and look at the public records and publications of Facebook researchers and academics working together in publications—there are a number of them.

Q1782 **Chair:** But it is interesting because in some ways people sought to portray you as some sort of rogue operator, but nevertheless you are someone who before the development of the app—we will come on to talk about that—was already working with Facebook and using Facebook data that had been given to you by Facebook.

**Aleksandr Kogan:** Yes, sir.

Q1783 **Chair:** Given some of the things that have been said about you, some people might find that quite surprising. Certainly, based on the evidence the Committee has received from Facebook, I think that is surprising. At the end of this particular pre-project that you were working on, what happened to the data that had been used to run these tests and analysis at the end of the process?

**Aleksandr Kogan:** We were going to hold on to it indefinitely, because we were writing many papers. Throughout that year—2013 and into 2014—we were writing about 10 different publications on it with different ideas. Normally, we would hold on to the dataset to generate further publications in the future. Once December 2015 comes along and

Facebook reaches out, they asked me to delete everything, including this academic dataset, which I did.

Q1784 **Chair:** Okay. So there was no requirement from Facebook, when they gave you that data in the first place, that you should destroy it or give it back.

**Aleksandr Kogan:** No. In fact, there was not even a signed agreement initially. They gave me the dataset without any agreement signed. It was just, "Here's an email. Here's the dataset." Sometime later—maybe even as far as a year later—we actually had a signed agreement. If you recall, there was a pretty big scandal about a "Facebook is trying to make people sad" publication—I think it was the fall of 2013—and in the wake of that they started to look more carefully at academic collaborations, and have them formalised a bit better.

Q1785 **Chair:** What was Facebook trying to achieve out of this, do you think?

**Aleksandr Kogan:** In terms of—

Q1786 **Chair:** What was the value to Facebook of you doing this work?

**Aleksandr Kogan:** I think it makes their employees happy. My perception was that management tolerated this. It was not a focus, because obviously it takes away time from their employees working on how to make Facebook a better platform. It was something they gave their employees to stimulate them—to allow them to have this relationship.

Q1787 **Chair:** Okay, so they are saying to their employees, "You can take macro-level Facebook user data and give it to academics who don't work for us without any kind of contract or licence, and let them play with it and just see what happens."

**Aleksandr Kogan:** Yes, sir.

Q1788 **Chair:** As simple as that.

**Aleksandr Kogan:** As simple as that.

Q1789 **Chair:** It seems pretty open, doesn't it?

**Aleksandr Kogan:** Yes. The company is very open.

Q1790 **Chair:** Maybe too open.

**Aleksandr Kogan:** Arguably so.

Q1791 **Chair:** Open not with their own information, but with their users' data.

**Aleksandr Kogan:** Given the last month, there is a lot of credibility to that statement.

**Chair:** Thank you very much.

Q1792 **Christian Matheson:** Dr Kogan, good morning. With the Chairman's permission, I might need to leave a little bit before the end of the evidence session, but no disrespect is meant.



## HOUSE OF COMMONS

You are an academic psychologist.

**Aleksandr Kogan:** Yes, sir.

Q1793 **Christian Matheson:** But you have also developed several apps that managed to harvest a large amount of data.

**Aleksandr Kogan:** Yes, sir.

Q1794 **Christian Matheson:** And that data then had a financial value of its own.

**Aleksandr Kogan:** Yes, sir.

Q1795 **Christian Matheson:** Which came first—the academic research or the idea that you might make a business out of this?

**Aleksandr Kogan:** Certainly the academic research.

Q1796 **Christian Matheson:** And at what point did you suddenly think, “Ah, there’s money to be made here”?

**Aleksandr Kogan:** If we stop and take a step back, the app was initially created to help us with the relationship we had with Facebook on the publications. As I said, we were writing 10 papers. I am a psychologist, so we were working at an individual level, typically, rather than the macro level that an economist or a sociologist would be interest in. We created this app to add data to the papers we were writing. Initially, it was that we would create the app, collect data at an individual level, and add it, with the Facebook data we have just discussed—

Q1797 **Christian Matheson:** Can I just check which app we are talking about at this point?

**Aleksandr Kogan:** This is the very app we have been dealing with throughout the story.

Q1798 **Christian Matheson:** And what is it called?

**Aleksandr Kogan:** It was called the CPW lab app. CPW is my lab’s name—the Cambridge Prosociality and Well-Being lab. I registered it under my personal account on Facebook—you have to do that as a developer. It was very simple to make. It is just a little bit of code, you get a log-in button, and then we would embed it in the studies we were running as a lab to collect the Facebook data. This was at some point in 2013. The conversation about a potential financial project came later, once I was introduced to SCL. This was spring 2014.

Q1799 **Christian Matheson:** So it was SCL that put the idea in your mind about maybe monetising the app.

**Aleksandr Kogan:** I think that is fair. It is slightly more nuanced, in that the conversation with SCL initially was not at all about Facebook. When I was first introduced to them, they just wanted consulting services and survey design. In our conversations over the next couple of months, we talked about other projects and other datasets. There was this interesting Facebook data that grew out of that.

Q1800 **Christian Matheson:** But you had, on a drive somewhere, a large bunch of data.

**Aleksandr Kogan:** Yes, sir. In my lab.

Q1801 **Christian Matheson:** Do you sell the data, or do you sell it as part of a package with your skills?

**Aleksandr Kogan:** We never sell that data. The data we collect as a lab is never sold.

Q1802 **Christian Matheson:** So is it transferred to—

**Aleksandr Kogan:** No. That would be a very serious breach of ethics. No. 1, that data never left my servers. It was used exclusively with my PhD students. No. 2, when we collected that data, we collected it by telling our participants it was for academic research, so we had no right to use it for anything else.

Q1803 **Christian Matheson:** So you would sell the app and somebody else would collect data using the app.

**Aleksandr Kogan:** Let me return to that in a second, so that it is clear—I know this has been a point of confusion. Thirdly, we did not collect names when we did the academic work. Even if we threw out the ethics and we threw out any concerns we did not even have the ability to do that. When we get to SCL, we collected new data and we used the app to collect new data and we stored it in a different server in a different location and under different terms of service.

Q1804 **Chair:** For clarity, I think Chris was asking about the CPW app, for which you are saying you held the data and it was never shared. Obviously, the My Digital Life app that we will talk about later, is separate from this.

**Aleksandr Kogan:** That is separate. It is not the app that was used for the project with SCL. Only a couple of hundred people ever used the thisisyourdigitallife app and that data was not given to SCL. The app was actually called the GSR app. Once we get to 2014 and we are going to do the project commercially, we changed the name of the GSR app and changed its description. Christopher Wylie provided us with commercial terms of service that we put on the Facebook platform. At that point, once we have made it commercial and separate from the university, we hook it up to a different dataset with a different database and different servers and the data goes there. So it is never in the same place. It can never be mixed up and new data is collected.

Q1805 **Christian Matheson:** Is the technical side of the app—the programming—the same?

**Aleksandr Kogan:** It is the same.

Q1806 **Christian Matheson:** You said you were introduced to SCL. One of the aspects of this inquiry that we have encountered is the interchangeability between SCL and Cambridge Analytica. You are clear you were working for SCL.

**Aleksandr Kogan:** I never had a contract with Cambridge Analytica.

Q1807 **Christian Matheson:** Did you ever have any contact with them?

**Aleksandr Kogan:** That is tricky in the sense that the folks who are involved in one are very much involved in the other. I worked with Alexander Nix and met him a couple of times. He is obviously involved with both. The data science team is very similar on both sides. Alexander Tayler is currently the head of research. I interacted with him. I interacted with all of them in their capacity as SCL Elections, but obviously they are now part of Cambridge Analytica and I am not even entirely clear what the distinction is.

Q1808 **Christian Matheson:** Were you clear at the time when you were working with them?

**Aleksandr Kogan:** I did not hear the name Cambridge Analytica when we started working with them. It did not come up. I only probably heard about it in January 2015, after we had already collected all the data.

Q1809 **Christian Matheson:** I want to ask you a question to which the specific answer is probably commercially confidential, but you can give us a ballpark figure. When you are selling an app to a company such as SCL, if I was to do the same thing at the same time, roughly how much might I be negotiating for? I am sure you do not want to give away—

**Aleksandr Kogan:** I can be fully transparent. To provide a bit of context: the app itself is not very expensive, because you can write this app in three or four days. It is not technically challenging in any way. Facebook explains how to do it, so there is great documentation on this.

Do not think of the app itself as something that is valuable, whereas the data that comes in is valuable. For the project that we did, the goal was to recruit about 200,000 people to authorise the app. There are various way to recruit that many people. There has been some confusion about this. Some folks believe there was a honeypot where we made a viral quiz, put it on Facebook, and people clicked—that is not the case. Those things are hard to do and it was not our avenue.

Instead, we basically did a market research project where we asked a company that specialises in recruiting people to recruit 200,000 people for us to do a survey about personality and various other items. As part of the survey there was a log in button that they clicked and that authorised the data. The money here, the expensive bit, is paying those users because each person costs about \$3 to \$4. Obviously, we were recruiting about 200,000 people so we were in that \$600,000 to \$800,000 range, just to pay the users. That was indeed what Cambridge, or SCL in this case, paid for during that first step of the project. Later on, they came to us and they wanted to get predictions. We use the data to drive predictions about people's personalities. For that, they paid us £230,000. I think that I testified to that in the written submission.

Q1810 **Christian Matheson:** That goes to you, your company or the university?

**Aleksandr Kogan:** The company.

Q1811 **Christian Matheson:** Did you have any day-to-day operational contact with them once you handed over the app?

**Aleksandr Kogan:** We ran the app. The way the project ran is that they would give me the survey questions, and I would be in charge of running the survey and collecting the data. This arrangement makes sense, because this way I control the data—I receive the data and have the data—and I can do the models on it and then deliver what we were contractually obligated to do.

Q1812 **Christian Matheson:** Just to be clear, unlike the academic data, this had names and all the identifying—

**Aleksandr Kogan:** Yes, sir.

Q1813 **Christian Matheson:** Can I ask one more question? Just to take you to the introduction with SCL, you suggested that Chris Wylie had given you the terms of service.

**Aleksandr Kogan:** Yes, sir.

Q1814 **Christian Matheson:** Did you deal directly with Mr Wylie?

**Aleksandr Kogan:** Yes, sir.

Q1815 **Christian Matheson:** Were there any lawyers involved? Did you have a lawyer?

**Aleksandr Kogan:** I did not. Chris Wylie said he was a data expert and an expert in data law. He guided us. This is well documented, and I can provide the Committee with evidence of that.

**Christian Matheson:** Thank you.

Q1816 **Ian C. Lucas:** When Facebook gave you information, could you identify individuals from that information?

**Aleksandr Kogan:** Do you mean the dataset they gave during the research project?

Q1817 **Ian C. Lucas:** Yes.

**Aleksandr Kogan:** No.

Q1818 **Ian C. Lucas:** Was there other data where you could identify individuals given to you by Facebook?

**Aleksandr Kogan:** Not through the API; directly by the company—is that correct? You mean not through the API and not through the app, but directly by the company?

Q1819 **Ian C. Lucas:** Yes.

**Aleksandr Kogan:** No.

Q1820 **Ian C. Lucas:** So at no stage did Facebook give you information that



## HOUSE OF COMMONS

could identify individuals.

**Aleksandr Kogan:** No.

Q1821 **Ian C. Lucas:** I want to ask you about your relationship with Cambridge University. I note that you are a research associate and university lecturer at the University of Cambridge in the department of psychology. You were appointed to that in 2012.

**Aleksandr Kogan:** August 2012.

Q1822 **Ian C. Lucas:** Do you receive a salary in that post?

**Aleksandr Kogan:** I do.

Q1823 **Ian C. Lucas:** Did you apply for the job?

**Aleksandr Kogan:** I did.

Q1824 **Ian C. Lucas:** Are you associated with an individual college?

**Aleksandr Kogan:** Yes, I am.

Q1825 **Ian C. Lucas:** Which college?

**Aleksandr Kogan:** Magdalene College.

Q1826 **Ian C. Lucas:** Do you teach individual students in tutorials and so on?

**Aleksandr Kogan:** I do.

Q1827 **Ian C. Lucas:** We have been talking also about the various companies that you have set up. There are a number of different companies.

**Aleksandr Kogan:** There are two.

Q1828 **Ian C. Lucas:** What are the names of those companies?

**Aleksandr Kogan:** There is Global Science Research, which is a UK entity, and then there is Philometrics, which was a US entity.

Q1829 **Ian C. Lucas:** When did you set up GSR?

**Aleksandr Kogan:** GSR was set up some time in the spring of 2014.

Q1830 **Ian C. Lucas:** And what about Philometrics?

**Aleksandr Kogan:** I believe we set up Philometrics in the summer of 2015.

Q1831 **Ian C. Lucas:** Does the intellectual property relating to those businesses belong to you, the companies or the university?

**Aleksandr Kogan:** They belong to me.

Q1832 **Ian C. Lucas:** They belong to you as an individual.

**Aleksandr Kogan:** They belong to the company, but the university makes no claim on the intellectual property that was brought over when I set up the company. My understanding in talking to the university is that since the only thing here that is valuable is code or know-how, the university



## HOUSE OF COMMONS

makes no claim on those things. It encourages its faculty members to start companies to commercialise their work.

Q1833 **Ian C. Lucas:** On the Cambridge University website, you are listed both as Aleksandr Kogan and Aleksandr Spectre. Can you explain that?

**Aleksandr Kogan:** Yes, sir. In 2015, I was married. My wife and I decided that it did not make sense for me to take her name or for her to take my name, so we said, "Let's choose a new last name." Since we are both religious and scientists, we thought the idea of light made a lot of sense. We were looking for something relating to light. My father was sadly sick at the time, and one of his surgeons was named Jason Spectre. We thought, "That is a really cool-sounding name", and it also nails down the theme of light because of spectrum. We decided on Spectre and a derivative of spectrum as a symbol of us going forward as a family.

Q1834 **Ian C. Lucas:** You know that Spectre is the evil organisation in the Bond films.

**Aleksandr Kogan:** It is an unfortunate coincidence.

Q1835 **Ian C. Lucas:** Did you know that at the time?

**Aleksandr Kogan:** I did not.

Q1836 **Ian C. Lucas:** You are not a James Bond fan.

Q1837 **Chair:** He's a Spectre fan, though.

**Aleksandr Kogan:** Exactly.

Q1838 **Ian C. Lucas:** Going back to the app you mentioned earlier, we were talking about individual data from Facebook. Can you identify individuals through the API?

**Aleksandr Kogan:** Yes sir. When you collect data through the API, a user authorises the app, so you can easily identify that user. Depending on what other data you gather, you could also identify their friends.

Q1839 **Ian C. Lucas:** You can identify their friends.

**Aleksandr Kogan:** Yes, sir.

Q1840 **Ian C. Lucas:** As well as the people who consented.

**Aleksandr Kogan:** Correct.

Q1841 **Ian C. Lucas:** Has that always been the case?

**Aleksandr Kogan:** Yes, sir.

Q1842 **Ian C. Lucas:** Did that change at all in 2014 with the terms of service from Facebook?

**Aleksandr Kogan:** What changed was that it was hard to not gather friend data. In terms of the user, you could still always identify them. On Facebook, your name, location, gender and birthday, I believe, are considered your public profile that anybody could search for.



## HOUSE OF COMMONS

I do not have Facebook anymore but if I did, I could search for you, sir. Then I could look at your profile and I would probably be able to see your name and some of this other information, even if we are not friends.

Q1843 **Ian C. Lucas:** I think what's difficult for people—well for me and, I think, others too—to understand is the distinction you draw between knowledge about individuals that is given to you by Facebook and knowledge that you derive through the API. Can you expand on that? What's the difference?

**Aleksandr Kogan:** The difference is that when you work with Facebook directly—with the datasets we got—they took millions of people and they averaged them together, so you lose the individuality and the differences. You just have a summary. That is what we are working with. From a summary, you can't scroll back.

For example, if we took the room here and averaged everybody's age together, and let's say the average was 45. Taking the number 45, I cannot know your age because you are just part of the group. That's the problem: there is no way to go back once you've summarised. Facebook gave me a summary. With the API you are getting your information, so I know exactly how old you are, if you have given your birthday.

Q1844 **Ian C. Lucas:** Essentially, you are deriving the same information: you are deriving personal information through the API?

**Aleksandr Kogan:** The API gives you the personal information; that is correct. You never get personal information from the data Facebook gave me through our academic link.

Q1845 **Ian C. Lucas:** So from my personal perspective, as an individual who is on Facebook, what really is the difference between personal information being handed over by Facebook, and you deriving that information from the API?

**Aleksandr Kogan:** They're just not giving us your personal information. They never say: "Ian Lucas, 40 years old". I never get that; I get: "England; average age is 42". At that point, there is just no real case for this being personal information because there are no people. It is a summary; it's a country at this point. Whereas, obviously, me getting information about you—and having a record in my database that says, "Ian Lucas, 42"—is a very different case.

Q1846 **Chair:** I'll ask a couple of questions about work at Cambridge University on this sort of analysis.

Other tools have been created at the university that help analyse Facebook likes and preferences to determine information about their data. I believe Cubeyou was one of those. Do you know much about the work of Cubeyou? Were you involved at all?

**Aleksandr Kogan:** I wasn't involved with Cubeyou. I know something about the collaborator they worked with, which is the Psychometrics Centre.

Q1847 **Chair:** On the website of the Psychometrics Centre, there is also an API called Apply Magic Sauce. Do you know who owns the licence for Apply Magic Sauce?

**Aleksandr Kogan:** I would believe the Psychometrics Centre.

Q1848 **Chair:** Do you know who owns the underlying data. Would that be owned by the Psychometrics Centre as well?

**Aleksandr Kogan:** I believe the myPersonality dataset was collected by David Stillwell before he joined the university. I believe he would himself own the underlying dataset. The tools I believe were built during his time in the university as part of the Psychometrics Centre, so I believe the centre would in theory—actually, let me back up. My understanding of the university's position is that the person who writes the code often owns it, unless it is contractually stated that they don't. I don't know who wrote that code and what contractual obligations and statements they have. It is possible they had a developer that did assign it to the university and the Psychometrics Centre. Or it is possible they still own the code itself in theory.

Q1849 **Chair:** With a project like that, who would benefit? Just the developer himself, or would it be more widely available to researchers at the university?

**Aleksandr Kogan:** The Psychometrics Centre is a bit of an unusual entity in the university, in that it is involved in both research and commercial activities under the same umbrella. That is more rare. I'd split mine up. I had my university position, which was very separate from my company activity. This was more under the same umbrella. My understanding is that the Psychometrics Centre uses its data for both academic work and commercial work and also just for fun—for people to get results about what their page likes say.

Q1850 **Chair:** Do you believe there was a commercial application from APIs like Apply Magic Sauce?

**Aleksandr Kogan:** I believe they certainly tried to find a commercial application. I am aware they were talking to many commercial entities. I just don't know where those conversations went.

Q1851 **Chair:** Okay. What sort of commercial application would that have been? Would that have been using the data that they gathered as a predictive model for targeting Facebook users?

**Aleksandr Kogan:** I don't know if it is quite that specific. My understanding of the way that they envisioned it is a company would bring in its own Facebook data, with its page likes, pass it through this Apply Magic Sauce API, get back predictions about various psychological traits, which then that company can use for whatever purpose its wants, such as market targeting or whatever else.

Q1852 **Brendan O'Hara:** Dr Kogan, could you tell the Committee under what circumstances you first met SCL?



## HOUSE OF COMMONS

**Aleksandr Kogan:** I was introduced to SCL by a PhD student at Cambridge. I knew the student because I had given him statistics advice on his PhD from time to time.

Q1853 **Brendan O'Hara:** What was the purpose of the meeting?

**Aleksandr Kogan:** My understanding was that the student was part-time consulting for SCL Elections and he had a friend there, named Chris Wylie, who he thought would be great for me to meet and see where the conversations go.

Q1854 **Brendan O'Hara:** When was that meeting?

**Aleksandr Kogan:** I believe it was January or February 2014.

Q1855 **Brendan O'Hara:** You met with SCL in January or February 2014.

**Aleksandr Kogan:** Correct, sir.

Q1856 **Brendan O'Hara:** And that meeting was initiated by someone working for SCL.

**Aleksandr Kogan:** Correct, sir.

Q1857 **Brendan O'Hara:** What was the outcome of that meeting?

**Aleksandr Kogan:** If my recollection is right, we talked about potentially collaborating and me providing them some consulting services. We also talked about them providing me with some of the commercial datasets they had acquired so I could set up a big data science institute.

Q1858 **Brendan O'Hara:** Was your decision to set up GSR directly related to your meeting with SCL?

**Aleksandr Kogan:** GSR was set up entirely to do the project with SCL—I actually relied on their guidance on how to do it appropriately—but it came a little bit later. At this point, we are not talking about commercial entities. Even the Facebook project is still a little bit off from coming up.

Q1859 **Brendan O'Hara:** Just for the record, you met with SCL in January 2014 and you set up GSR in May 2014, specifically to work for SCL.

**Aleksandr Kogan:** Correct, sir.

**Brendan O'Hara:** In June 2014, the director of analytics at SCL wrote to Cambridge University to say that Dr Kogan first introduced them to the possibility of using online social media data to score and predict human personality traits at a meeting in London, as we have said, in January 2014. Upon agreeing to explore the viability of that idea, Dr Kogan then introduced them to Dr David Stillwell at a meeting in Cambridge in February to discuss the myPersonality project and whether it could be used in partnership with SCL. Why would the director of analytics at SCL write to Cambridge University?

**Aleksandr Kogan:** Initially, when the project was envisioned, there was going to be a collaboration between myself and the Psychometrics Centre. The idea was that since the Psychometrics Centre had already set up an



## HOUSE OF COMMONS

API and models for personality, we would just use that to generate the predictions. On my side, I would use the app that I had built to collect the data. So, at some point during these negotiations, SCL floated a rough budget of about \$2 million. I think it was something that Chris Wylie mentioned during the conversation.

I thought, all right, maybe we will give the Psychometrics Centre \$500,000 for the modelling aspect of this. As I just mentioned, I was trying to set up a big data science institute—that was my intention. So there was a lot of money—I think a couple of hundred thousand dollars—allocated for server costs and running that.

The Psychometrics Centre was fine with that and they liked the idea. As the conversations continued, SCL backed off; they really clawed. They said, “We are not going to pay for anything other than data collection at this point.” That was the \$800,000 to pay the participants to answer the survey.

They said, “We don’t see why we should be paying \$500,000 for the models that the Psychometrics Centre is providing.” At that point, I went to the Psychometrics Centre and said, “They are not willing to pay it. How about \$100,000?” I was trying to broker a middle road. The Psychometrics Centre refused. So, SCL instructed me to remove them from the project.

At that point, Dr Stillwell and Dr Kosinski, the two people at the Psychometrics Centre I was talking to about this, went and informed Professor Rust of the situation, because he had never been involved with this. Professor Rust wrote a complaint letter to the university saying that “I think Alex is trying to swindle us out of this. He is going to get \$1 million. We were promised \$500,000 and now there is going to be \$100,000.” So, he made some serious accusations. SCL wrote a letter to the university in my defence to state that that was just not true. The Committee has already seen the contract that we ultimately signed with SCL, so you also know that is not true.

**Q1860 Brendan O'Hara:** Okay. So the timeline was, you met SCL in January 2014. They then engaged Cambridge University in February 2014. Cambridge University decided not to go ahead with it for whatever reason. Then you set up two months later GSR.

**Aleksandr Kogan:** That’s not quite right. The problem with that account is that it is not Cambridge University. It is this one lab at the Psychometrics Centre.

**Q1861 Brendan O'Hara:** But the director of SCL in his email wrote to Cambridge University. I am trying to work out who he spoke to, why he spoke to Cambridge University, then for nothing to come of it, and then for you to set up within weeks GSR.

**Aleksandr Kogan:** The plan was already to set up GSR. That was not something new. The intention throughout this process was to set up GSR. Because there was going to be this collaboration between GSR—a private entity that I held—and the Psychometrics Centre, which is a lab in the

university, which as I said is a bit unusual in that it works both in a commercial and academic space, so we were going to collaborate.

The breakdown occurred because the Psychometrics Centre refused to move from this price tag of \$500,000, so SCL kicked them off the project. At that point, the Psychometrics Centre made a complaint to the university, to their legal team, "Hey, we think this is inappropriate and we were promised \$500,000." So SCL wrote to the university legal team to explain that what the Psychometrics Centre had alleged was untrue.

**Q1862 Brendan O'Hara:** But that is not quite the account in this email. As I said, the then director of SCL in February wanted to discuss the myPersonality project and whether it could be used in partnership with SCL, having already looked at online social media data to score and predict human personality traits. I am struggling to see how one went from this meeting—what you described as a fairly informal meeting—with an associate of SCL in the January, all the way through February with this great plan that never came to fruition, then you setting up GSR and working with SCL.

**Aleksandr Kogan:** Let me walk you through that in a bit more detail. I met SCL—Chris Wylie in particular—and we started these conversations. They were trying to pull out many datasets. They were buying up many datasets from data brokers—at least, that is what I was led to believe. I said, "Well, listen, if you are interested in datasets, and I know you are interested in personality"—the interest in personality was already there—"let me introduce you to the Psychometrics Centre, who hold a very big dataset on personality." That was the myPersonality project. I facilitated a meeting for them to strike that conversation. SCL are at this point interested in acquiring this entire dataset, but David Stillwell, who owns the dataset as we have already discussed, decides not to sell it to them. The grounds were that he had collected that data by telling people it was going to be for academic work, so he felt it would be inappropriate to sell it.

All fine, so far. At this point, David, Michal Kosinski—who is also part of the Psychometrics Centre—and I propose an alternative plan. Rather than using the dataset they already had, we would use just the models they had built to make predictions, but we still need data to feed through these models. So I would collect the data, through GSR—that was always the plan—and feed it through these models that the Psychometrics Centre own, which is the Apply Magic Sauce, and generate these predictions about personality. The bit that never came to fruition was this passing the data through their models, because, like I said, they wanted half a million dollars. The only thing that changed was that GSR was not only responsible for the data collection but also for the modelling aspect and the predictions.

**Q1863 Brendan O'Hara:** Okay. So you then set up GSR on your own.

**Aleksandr Kogan:** With one of my associates.

**Q1864 Brendan O'Hara:** So there are two of you. You are the two company

directors of GSR.

**Aleksandr Kogan:** Yes, sir.

Q1865 **Brendan O'Hara:** And you have this relationship with SCL. Did SCL pay you for your services?

**Aleksandr Kogan:** At this point they paid only for the data collection. Later on, in January 2015, when we delivered them the second tranche of data, they paid GSR £230,000.

Q1866 **Brendan O'Hara:** How much did they pay initially, before the payment of £230,000 in January 2015? How much were you paid by SCL or any of their associates then?

**Aleksandr Kogan:** In terms of me, or my company?

Q1867 **Brendan O'Hara:** The company.

**Aleksandr Kogan:** We netted zero. The money that went through—the £800,000 I think you are referring to—is the cost for the participants. That is paid almost entirely directly to Qualtrics. That money did not even pass through us. So we would get an invoice from Qualtrics and give it to SCL, and SCL would pay that invoice directly. SCL gave us maybe £10,000 or £20,000, if I recall correctly, to pay for the servers we were running on Amazon Web Services. All that money was entirely for that and we had to match anything we spent with an invoice. So the company at that point was not paid any money at all.

Q1868 **Brendan O'Hara:** The only money you made as a company came in 2015, and that was a £230,000 payment from SCL.

**Aleksandr Kogan:** Correct, sir.

Q1869 **Brendan O'Hara:** Was your contract exclusive to SCL or could you approach, or were you approached by, any other data analytics companies?

**Aleksandr Kogan:** It was not exclusive. The way my contract was set up was that I owned the dataset we collected. We were approached by a company called Eunoia, again in summer 2014. This was a company that Chris Wylie set up immediately after leaving SCL.

Q1870 **Brendan O'Hara:** So when you were approached by Christopher Wylie, did you work with Christopher Wylie, did you sell Christopher Wylie or Eunoia any of your data?

**Aleksandr Kogan:** We did, but it was not for a monetary exchange. The agreement was that we would give them our dataset in exchange for the datasets he purported to have. He did not honour that agreement. He never delivered his datasets, so we moved to cancel that contract by a legal letter in, I believe, March or February of 2015.

Q1871 **Brendan O'Hara:** Just bringing this to an end, Christopher Wylie maintained that he refused to work with you when he realised you were not setting up an academic institute as he believed you had originally



## HOUSE OF COMMONS

claimed, but that you were setting up this private company with Alexander Nix. How do you respond to that claim?

**Aleksandr Kogan:** That is a fabrication.

Q1872 **Brendan O'Hara:** In what way?

**Aleksandr Kogan:** In all ways. Chris Wylie helped me to set up the for-profit entity. It was literally he who said to me, "Please set up a for-profit entity". Then, after he left SCL, he tried to get me to work with his company. I was not interested, after I had witnessed how he interacted with SCL. He still wanted to have a partnership, so we agreed to do a fair commercial agreement. I gave his company the dataset. The only time this broke down was when he did not honour his side of the agreement and did not deliver the data and we moved to cancel it in February or March 2015. Up to that point, I would say that Mr Wylie was highly enthusiastic about working with me and my commercial entity.

Q1873 **Brendan O'Hara:** What was your personal relationship with SCL, and were you paid personally by SCL ever?

**Aleksandr Kogan:** Never.

Q1874 **Brendan O'Hara:** Or any of their associates?

**Aleksandr Kogan:** Never.

Q1875 **Brendan O'Hara:** Okay. What is the status of GSR now?

**Aleksandr Kogan:** It has been closed.

Q1876 **Brendan O'Hara:** When was it closed, and why?

**Aleksandr Kogan:** I believe it was closed maybe a year ago. I would have to double check the date for you. It was closed because—  
[Interruption.]

Q1877 **Brendan O'Hara:** From my reading the company was dissolved in October 2017. Why was it dissolved?

**Aleksandr Kogan:** Initially when I set up the company, I was not really thinking about making a company. I was really interested in setting up a non-profit big data institute. I wanted to call it "Mutual Science". Mr Wylie advised me, "Oh you should also set up a for-profit entity." I think the reason was that there would be tax benefits. That was Global Science Research.

The initial plan was that we would collect the data, I would fulfil my obligations to SCL, and then I would use the data for research. After the project, we decided to give the whole company thing a go. Another associate joined the company who actually had some business experience, because I had none, to try to make a company of it. We tried to develop a couple of products—one was brand reports—but it was just never successful. We used the £230,000 that SCL eventually paid us to invest in trying to build this product. We bought some Twitter data, hired some



## HOUSE OF COMMONS

developers, and tried to sell it, but we never got a single client, so the money ran out. We closed the company just because it failed.

Q1878 **Ian C. Lucas:** Why was GSR registered at the 29 Harley Street address?

**Aleksandr Kogan:** Is that in Cambridge?

Q1879 **Ian C. Lucas:** I think it is in London. Do you know what the registered office of GSR was?

**Aleksandr Kogan:** Initially it was just my apartment in Cambridge. I think later we moved it to an innovation centre in Cambridge, and later Manchester. I am legitimately surprised by this.

Q1880 **Ian C. Lucas:** Did you use an agent to set it up?

**Aleksandr Kogan:** We used Formations House.

Q1881 **Ian C. Lucas:** Who is we?

**Aleksandr Kogan:** Myself and my associate who I made the company with.

Q1882 **Ian C. Lucas:** Who was that?

**Aleksandr Kogan:** I want to keep the name private—again, in confidence.

Q1883 **Ian C. Lucas:** Are they on the company's register?

**Aleksandr Kogan:** They are.

Q1884 **Ian C. Lucas:** Was that Joseph Chancellor?

**Aleksandr Kogan:** It was.

Q1885 **Ian C. Lucas:** So you don't know anything about the Harley Street address?

**Aleksandr Kogan:** I am legitimately surprised by that.

Q1886 **Ian C. Lucas:** It is used by a lot of shell companies, some of which have been used for money laundering for Russian oligarchs. Did you know that?

**Aleksandr Kogan:** I did not. I'm unfortunately not a Russian oligarch.

Q1887 **Ian C. Lucas:** Have you met any Russian oligarchs?

**Aleksandr Kogan:** I have not.

Q1888 **Ian C. Lucas:** Did SCL ever tell you that they would be using the information that they were receiving in connection with these projects for political campaigning?

**Aleksandr Kogan:** The official name of the company was SCL Elections, so it would be hard to miss that.

Q1889 **Ian C. Lucas:** So you knew it was going to be used for political



## HOUSE OF COMMONS

campaigning.

**Aleksandr Kogan:** Yes, sir.

Q1890 **Ian C. Lucas:** Did they specify where it was going to be used for political campaigning—which countries? *[Interruption.]*

**Aleksandr Kogan:** What is that?

**Ian C. Lucas:** It is the start of business. When it is really exciting, it is a vote, but it is not a vote at the moment. It is just the start of business.

**Aleksandr Kogan:** Okay. It was going to be in the United States, given that we were collecting data for the United States.

Q1891 **Ian C. Lucas:** Was there any discussion about political campaigning in the UK?

**Aleksandr Kogan:** Later. After we had done this project, there was some discussion about it, but it never went anywhere.

Q1892 **Ian C. Lucas:** Was that discussion with Alexander Nix at any time?

**Aleksandr Kogan:** Yes, it was.

Q1893 **Ian C. Lucas:** So Alexander Nix talked to you about political campaigning in the UK?

**Aleksandr Kogan:** He did.

Q1894 **Chair:** Obviously Joseph Chancellor's name has now come up. Can you explain his role as a director in GSR? Was he your partner? Did you work on everything together?

**Aleksandr Kogan:** We did. When the company was set up, we were 50:50 partners. Eventually another co-founder came in, so we became equal third partners.

Q1895 **Chair:** So the work you were doing for SCL, the work you were doing looking at Facebook data, he was your partner across all of that?

**Aleksandr Kogan:** Yes, sir.

Q1896 **Chair:** He has worked for Facebook since November 2015, hasn't he?

**Aleksandr Kogan:** Yes, he has.

Q1897 **Chair:** So when Facebook's response from their deputy general counsel, describing your work as a "scam" and a "fraud", as data-harvesting—and they singled you out to say, you have lied to us and violated our platform policies, those remarks must apply to Joseph Chancellor as well.

**Aleksandr Kogan:** If you want to push on the spirit, I would agree. Personally I am very glad that they have not moved on Joe—I think it would be petty, personally.

Q1898 **Chair:** But you say you work on it together, so if it applies to you, it applies to him in the same way.

**Aleksandr Kogan:** I think that would be a fair characterisation.

Q1899 **Chair:** So do you not think it is odd that they employ someone that they regard as a scam and a fraud, and who has lied to them and violated their platform policies?

**Aleksandr Kogan:** Honestly, I don't think it is odd.

Q1900 **Chair:** But I think any normal person would think it is odd. It may not be odd in Facebook-land, but—

**Aleksandr Kogan:** The reason I don't think it is odd is because, in my view, Facebook's comments are PR-crisis mode. I don't believe they actually think these things, because I think they realise that the platform has been mined left and right by thousands of others. I was just the unlucky person that ended up somehow linked to the Trump campaign, and we are where we are. I think they realise all this, but PR is PR, and they are trying to manage the crisis. It is convenient to point the finger at a single entity and try to paint the picture that this is a rogue agent.

Q1901 **Chair:** But do you not think it is surprising that Facebook employs someone who knew all about this, all about what was being done—that this Facebook data was being gathered to be used by SCL in elections in America—yet did very little to investigate that, even after the story was made public by *The Guardian* newspaper at the end of 2015, when Joseph Chancellor was already working for the company?

**Aleksandr Kogan:** Yes, sir. In fairness to Facebook, my perception is that they did a pretty thorough job of investigating the issue. They talked to all the parties involved; they demanded that all the data be deleted. The difficulty here is compliance. I have seen that Facebook has been challenged—as in, why didn't you audit anyone? The problem with an audit is that the only thing you are going to catch is people who are trying to do the right thing, but missed a few files—because that is possible. You're trying to delete everything, but maybe you missed a file here and there, because you don't have forensic resources. What an audit cannot do is catch people who are trying to be bad actors, because you can always put the data on a hard drive and stick it under your mattress. So in many ways that's a futile effort. Once the data is off the system, you really are on the honour system in terms of ever trying to put that genie back in the bottle.

Q1902 **Chair:** Sandy Parakilas gave evidence to the Committee a few weeks ago. He basically said that once the data is gone, it's gone. Even if the original dataset is destroyed, the derived value of the data can never be recovered, because it then exists in a new form. Would you agree with that?

**Aleksandr Kogan:** Facebook asked everybody to delete the data and its derivatives. So I think if you are careful in scope and really make sure that everything that could be considered derivative is deleted, you could delete that. The problem is that you are still on the honour system, because if somebody decides not to—and I think in this case there is now evidence to suggest that somebody decided not to be honourable in terms of all the



## HOUSE OF COMMONS

parties involved, and didn't delete the data—it is very hard to know that until there is a PR crisis or somebody sees that and they get a tip.

Q1903 **Chair:** Yes. But, certainly, the data you acquired was used during the Ted Cruz campaign, wasn't it?

Q1904 **Aleksandr Kogan:** I believe so, yes. SCL admitted as much to me.

Q1905 **Chair:** Do you believe that the derived data was also used in the Trump campaign in part as well?

**Aleksandr Kogan:** I can't know.

Q1906 **Chair:** But do you believe it would be likely?

Q1907 **Aleksandr Kogan:** I would actually think it would be unlikely, personally.

Q1908 **Chair:** Given that they have already said that they used data from the Cruz campaign in the Trump campaign?

**Aleksandr Kogan:** The reason I think it is unlikely is that, from talking to others who knew a bit more—because I was not involved at this point; this is much after my time—my understanding is that the Ted Cruz campaign was part unhappy with the product that they thought that SCL delivered to them. Also, SCL at this point is under enormous pressure from Facebook to clean up their act and delete all the data.

If I am a company, I don't know that I would risk the legal liability of somebody like Facebook suing me to keep a dataset that has apparently failed with the last client that it was used for. I don't know that this was their thinking, but it does not necessarily make sense to me that they would keep this around.

Q1909 **Chair:** Can I just ask a couple more questions on this before bringing in colleagues: the political work that SCL was engaged in, which you were supporting—this included the work for Ambassador Bolton's super PAC, didn't it?

**Aleksandr Kogan:** That I don't know.

Q1910 **Chair:** Because there are emails that we published, given to us by Chris Wylie, which reference you working on gathering the data for that work. So you were gathering data on state-specific work for voters in Arkansas, North Carolina and New Hampshire to support Ambassador Bolton's political activities in those states—the campaigns he was supporting.

**Aleksandr Kogan:** I would say that the details of who the candidates were was obfuscated for me. What I was told was, "We need these 11 states," and I was given the surveys that they wanted to ask in the collection. In terms of the details of who they were working with, that was not really ever told to me. Ambassador Bolton is a name I have recently heard. It was not a name that I can ever recall hearing at that time.

Q1911 **Chair:** In this email it says—this email was sent by someone on the SCL team—"I also recall your very good idea that Kogan could model 2/3 of



## HOUSE OF COMMONS

what he had of that last round of Bolton issues testing, thereby enabling faster further modelling by the data team to be done first on that sample, then the final model once all data was received. Kogan gets data daily so 2/3 of the data must already be available, even with the delay due to Qualtrics's mistake. Was this done and can the preliminary findings be shared"? This sounds like a fairly dynamic process that you are working on. It is feeding directly into the work for the Bolton super PAC.

**Aleksandr Kogan:** It is less dynamic than it sounds, because the way the project ran is we did a few waves of data collection and for each wave of data collection SCL would provide me "This is the survey we want to ask". I wasn't instructed about why it was being asked, just "This is the survey we want to ask." Then we would run the study, collect the data, and then we would apply the models we were building to generate predictions about each of these questions. The modelling process does not really even require us to know what the question is. It is just a number. We are just predicting that.

Q1912 **Chair:** So what you described earlier on—the survey is being set up; you have got a company going out, getting the bulk audience you need. To some people, when they heard you talk about it earlier, that may sound like a reasonably indiscriminate process, but this would suggest that people are being deliberately targeted in certain states, because that is useful to the commercial project that SCL are working on. So are you directing the surveys team and saying, "We need more people from New Hampshire to respond to this survey, to add into the model"?

**Aleksandr Kogan:** It wasn't quite that. We actually did collect data from all 50 states. It was closer to indiscriminate, because the goal was to make predictions on as many people as we could. In this case, with those 11 states, I think we delivered a couple of million people—2 or 3—but the thesis was, rather than a story just focusing on those states, it is better to go in all states, to get the data faster, because, let's imagine you live in Virginia; you might have friends in West Virginia, and so we would have got that as well. So we got the data, actually, across America. I don't recall we ever actually, when we instructed Qualtrics to collect the data, there were any restrictions.

Q1913 **Chair:** You can take a moment to read the notes, if that would be helpful.

**Aleksandr Kogan:** Just to make something else clear, the emails that you are talking about I was never involved with. I saw those emails actually for the first time yesterday when it was pointed out that this information has been submitted.

Q1914 **Chair:** Yes, but you are clearly interacting with the people that are in the email chain, because you are working for them.

**Aleksandr Kogan:** Yes—so I am chatting with some of the SCL team. If you can imagine, SCL is in between here, and I guess the Bolton super PAC is on one side instructing them. SCL is instructing me, but there is no communication between these two entities, and quite honestly not even awareness of this side.

Q1915 **Chair:** From our point of view, that doesn't really matter to us, whether you had direct interface with the Bolton team, because this is clearly a project that SCL are undertaking on his behalf, and you are clearly playing a vital role because you are gathering data that has been inputted into the tools that are used to support that campaign. So presumably you must have known or been told that there was a priority around processing data linked to certain states, because the Bolton super PAC project was focused at that time on specific races in specific states, not looking to build up a survey of the whole country—not at that point, anyway.

**Aleksandr Kogan:** We were told that we were focusing on specific states, but why we were focusing on those states, I have no recollection of ever being informed.

Q1916 **Chair:** But you said earlier that you knew it was elections.

**Aleksandr Kogan:** It was elections. Certainly it was elections in the nine or 11 states that we were contractually obligated to seek out.

Q1917 **Chair:** Okay. You did not necessarily know who the candidates were. You just knew that those were the state races you were working on.

**Aleksandr Kogan:** Exactly.

Q1918 **Chair:** Finally, I want to cover some of the evidence that Alexander Nix gave to the Committee. Can I read to you the questions I asked Alexander Nix and the answers he gave, and can you tell me whether you agree with the answers he gave?

**Aleksandr Kogan:** Sure.

Q1919 **Chair:** I said to him, "Does any of your data come from Global Science Research?", and he said no.

**Aleksandr Kogan:** That is a fabrication. Maybe not now—

Q1920 **Chair:** I said, "They have not supplied you with data or information?", and he said no.

**Aleksandr Kogan:** Total fabrication.

Q1921 **Chair:** I said, "Your datasets are not based with information you have received from them?"

**Aleksandr Kogan:** That could be true, depending on what the datasets are now.

Q1922 **Chair:** But previous datasets would have been?

**Aleksandr Kogan:** Sure. We certainly gave them data. That is indisputable.

Q1923 **Chair:** So as far as you are concerned, he lied.

**Aleksandr Kogan:** Absolutely.

**Chair:** Thank you.

Q1924 **Paul Farrelly:** I have a few different questions, but I want to close off some of the issues that have just been raised. If you were doing this work, why would you not want to know who the ultimate client was, particularly if it was elections?

**Aleksandr Kogan:** That is a good question. I knew it was for Republicans, so I think that is true, but honestly, I would not be able to recognise the names. I don't know the specific candidates in most primaries or elections in the United States. It is just not something—that level of granularity is not something that I have ever been interested in.

Q1925 **Paul Farrelly:** So you did not care, really?

**Aleksandr Kogan:** In terms of the specific candidates, no. That's—

Q1926 **Paul Farrelly:** There are some nice politicians around and some really nasty politicians, so why would you not care?

**Aleksandr Kogan:** About whether the politician was nice or not?

**Paul Farrelly:** About who the client was, yes.

**Aleksandr Kogan:** My personal position on life is that unless I have a lot of evidence, the answer is I don't know. That is a good lesson to learn from science, where typically we just don't know. In terms of politics in particular, I rarely have a very strong opinion about a candidate, whether they are nice or not. My general perception is that, especially in the United States, most folks are trying to do what they believe is right. For most major candidates and major parties, I can understand where they are coming from, even if I would not personally agree with that position. There are exceptions, obviously, but by and large, I would say that most candidates are coming from a position that could be viewed as reasonable by a substantial proportion of the population.

Q1927 **Paul Farrelly:** Surely another lesson from science is that if you don't ask the question, you don't get the evidence.

**Aleksandr Kogan:** Not quite. You have to be careful not to make a judgment too quickly. Something like, "Is this person a nasty person or a nice person?" is quite a complex question. It is difficult for me to know that about a politician when we are getting very select slices. I do look at the positions, and there will be positions I agree with and positions I personally disagree with—

**Paul Farrelly:** But you didn't bother.

**Aleksandr Kogan:** I didn't bother, no.

Q1928 **Paul Farrelly:** Joe Chancellor, 50:50, what did he bring to the party?

**Aleksandr Kogan:** He and I have very similar skill sets, in that we can both programme and we are both good at statistics and machine learning. We basically divided and conquered the tasks. Initially, my job was to run the app, because I wrote the code for the app to collect the data, and he was involved in doing all the modelling. I did not really do any of the

machine learning and modelling until probably the second month of the project.

**Q1929 Paul Farrelly:** So actually, Cambridge Analytica and SCL did not need the psychometrics centre, did they?

**Aleksandr Kogan:** No. The mistake the psychometrics centre made in the negotiation was that it believed that models are useful, rather than data, and it is actually the opposite. Data is far more valuable than models, because if you have the data, it is very easy to build models—you just use a few well-understood statistical techniques to make them. I was able to go from not doing machine learning to knowing what I needed to know in one week. That is all it takes, if you are competent in statistics.

**Q1930 Paul Farrelly:** Once SCL and Cambridge Analytica know that one of your survey people has liked, for example, an imaginary Facebook page out of the 500 called “I hate Hillary Clinton”, that is job done, isn’t it?

**Aleksandr Kogan:** Not quite.

**Paul Farrelly:** Almost.

**Aleksandr Kogan:** Remember that, when we built the models, we had all the Facebook page likes. We gave SCL a tiny portion of that. On Facebook there are more than 150 million pages that people can like. You are referring to our eventually giving them some page likes, but initially we kept it all and we built models on all of them. There is some knowledge and expertise required to understand how to do that correctly, but it is very noisy.

**Q1931 Paul Farrelly:** Mr Chancellor joined Facebook in 2015. How did you come to work together, and how did he come to, presumably, take an offer from Facebook that he couldn’t refuse?

**Aleksandr Kogan:** He came to me as a post-doc to my lab. My previous post-doc was actually hired by Facebook, so we needed to replace him. We were recruiting for the position, we published it publicly and Joe applied. My previous post-doc actually had a relationship with Joe—they had met before—so he highly recommended Joe. He came to work with me in that way. Initially it was just research in the lab, and then we decided to go into partnership, in terms of the business.

The Facebook side came about because Joe had another baby—he was already a father and he had a second child—and he needed a real income. Obviously, start-ups are very risky, so he decided to go and try to find a tech job. We had a good relationship with Facebook. We recommended Joe to Facebook, he interviewed and it went well.

**Q1932 Paul Farrelly:** I have a final question on loose ends. GSR has been wound up.

**Aleksandr Kogan:** Yes, sir.

**Q1933 Paul Farrelly:** You didn’t take a salary from it?

**Aleksandr Kogan:** Never.

Q1934 **Paul Farrelly:** Did it cover many of your expenses?

**Aleksandr Kogan:** It covered a few expenses, such as travel a couple of times, but it was negligible.

Q1935 **Paul Farrelly:** When it was wound up, was there any surplus left to be distributed?

**Aleksandr Kogan:** There was. When we started GSR, we put in £9,000. When we rounded up, I think it was around £8,000.

Q1936 **Paul Farrelly:** But none of the £230,000 was left? What happened to that?

**Aleksandr Kogan:** It was spent, first, on developers trying to build that brand product, which failed. It was also spent on buying Twitter data before that product, and it was then spent on lawyers negotiating with Facebook in the wake of December 2015.

Q1937 **Paul Farrelly:** With regard to your app, which you wrote, one of the neat tricks in the way it works is that it could get into people's networks and get their friends.

**Aleksandr Kogan:** Yes.

Q1938 **Paul Farrelly:** You pay a couple of dollars and, depending on how many friends someone has, you get a thousand for the price of one, almost.

**Aleksandr Kogan:** I think the average was closer to between 200 to 300 people, but yes.

**Paul Farrelly:** That was the smart mathematical trick, in terms of numbers.

**Aleksandr Kogan:** It was the efficient way of gathering data on Facebook at the time.

Q1939 **Paul Farrelly:** Could you just explain this for people? I only use one Facebook app. I play Scrabble; I spend far too much time playing Scrabble on Facebook.

**Aleksandr Kogan:** It's a great app.

Q1940 **Paul Farrelly:** On my Facebook, I have it set so that only my friends can see my friends. If I were to do one of your surveys, I have not become a friend of yours, so how does it actually technically work? You go from the survey to seeing my friends, presumably, if I am on that setting? It is not just that only I can see my friends.

**Aleksandr Kogan:** Sure. From 2006 to 2015, one of the core features of the Facebook API was that you could gather data about users and their friends. I think the thesis was that the data is highly innocuous and is basically public—all of your friends can see it.



## HOUSE OF COMMONS

The information that we could gather through the API about a user's friends was only things that the user and the rest of their friends could see. For example, your wall post on Facebook is something that anybody who is a friend of yours can see. However, your private messages are something that only you and the person you are talking to can see. The first class of data you could share, but the second class you could not, because that is not information I would have.

So long as, first, it was something visible to me as your friend, and secondly that your security settings permitted me, as your friend, I could share that data with the collective. It wasn't really a trick: it was a core feature of the system at that time.

Q1941 **Paul Farrelly:** It was central to what you were doing.

**Aleksandr Kogan:** It was central to what we were doing and to what a lot of companies were doing.

Q1942 **Paul Farrelly:** Sure. I will come on to terms and conditions in a moment. When I am getting a couple of dollars for a server, I am not aware that I am effectively making you one of my friends, and that you can therefore see all my friends. It could come as a bit of a surprise.

**Aleksandr Kogan:** It is not quite becoming one of your friends. You are granting me access to data about your friends.

Q1943 **Paul Farrelly:** Which is the same thing, really.

**Aleksandr Kogan:** On the Facebook platform there is a distinction. The other thing is you would be aware of the data we are gathering. When you log in and it says, "GSR app would like to collect some of your data," it tells you explicitly the things we want to collect. Facebook controls this workflow. They will say, "We want to get your name, your location and your page likes, and we want to get the same information for your friends." That would be quite front and centre in terms of the type of data being collected.

Q1944 **Paul Farrelly:** Looking at the terms and conditions of GSR and thisisyourdigitallife, your second two apps, which were used commercially, there is Magdalene College, which you mentioned before, and St John's Innovation Centre. Are they are both University of Cambridge establishments?

**Aleksandr Kogan:** Magdalene College was just where I lived. St John's Innovation Centre is a commercial business centre. We have a mailbox there. Initially, when I set up the company, I registered it literally where I lived, and then we moved to be at the innovation centre once we had our new business partner, so he could get mail from there.

Q1945 **Paul Farrelly:** But it is undeniable that Magdalene College is a Cambridge University—

**Aleksandr Kogan:** Yes, sir.

Q1946 **Paul Farrelly:** What sort of approvals do you have to get from the university when you register companies on their premises?

**Aleksandr Kogan:** I did not get any approval.

Q1947 **Paul Farrelly:** Does Cambridge insist on approvals for companies—

**Aleksandr Kogan:** Not that I am aware of.

Q1948 **Paul Farrelly:** So if no approvals were needed, presumably the university was not insisting on being quite clear about what companies registered on its premises were doing.

**Aleksandr Kogan:** That would be an overstatement, in my view. Just because a company is registered does not mean you have a crystal ball into what that company is doing in terms of activity.

Q1949 **Paul Farrelly:** No, the point I am making is that if you did not go through any approval process—if the university does not have one or, if it does, you did not go through it—there was no opportunity to ask questions about what the company was doing or whether the university approved of a company that was doing a certain thing being registered on its premises.

**Aleksandr Kogan:** Ah, yes, I would agree with that.

Q1950 **Paul Farrelly:** I will go back to this email before moving on. In terms of the payments, is there any reason why, in this email to you, John Rust, the head of Cambridge University Psychometrics Centre, would describe what you were planning as a “get rich quick” scheme at their expense?

**Aleksandr Kogan:** Paranoia.

Q1951 **Paul Farrelly:** He got the wrong end of the stick.

**Aleksandr Kogan:** Quite frankly, they are not here with me today because of greed. It is ironic. As I was describing before, initially Mr Wylie floated a budget of \$2 million. That eventually got clawed back, and John just did not believe it got clawed back. We now know that it did get clawed back, because you have the contract in hand.

Q1952 **Paul Farrelly:** Essentially, the payment, apart from the £230,000 for your role in this, was to keep the data.

**Aleksandr Kogan:** The £230,000 came as a second agreement later.

Q1953 **Paul Farrelly:** Yes, but essentially your reward for doing this was to keep the data.

**Aleksandr Kogan:** Yes. Exactly.

Q1954 **Paul Farrelly:** Which you could then use in your academic life or—

**Aleksandr Kogan:** Correct.

Q1955 **Paul Farrelly:** That was part of the deal.

**Aleksandr Kogan:** That was the deal. I was rewarded with data.

Q1956 **Paul Farrelly:** You say in your evidence that you have had university ethics approvals for all your academic work.

**Aleksandr Kogan:** Yes, sir.

Q1957 **Paul Farrelly:** Did you have university approval for that deal?

**Aleksandr Kogan:** For the commercial activities?

Q1958 **Paul Farrelly:** Yes.

**Aleksandr Kogan:** There is no real mechanism for a company to go out and seek ethics approval for a commercial deal. The university oversight is over academic activities. This fell outside that.

Q1959 **Paul Farrelly:** But this was a deal in which your reward was data that might be used in academic work.

**Aleksandr Kogan:** It could be. The process normally in this sort of situation is that if commercial activity generates a dataset and you want to bring it into the university, you take that dataset and you bring it into the university through a data transfer agreement, and we were in the process of doing that afterwards and working through some of the issues with the university. But I have never heard of anybody who runs a company trying to get ethics approval for a dataset whose primary function was really a commercial enterprise. Our primary deliverable here, first and foremost, was the obligations we had to SCL. Secondary purposes come later, when you try to bring the work in for the university.

Q1960 **Paul Farrelly:** Perhaps these are aspects that the university might want to look at, if it does not already. I do not want to take too much more time. Can I come to your terms and conditions? On the GSR, clause 3 on the purpose of the application states, "We use this application as part of our research in understanding how people's Facebook data can predict distant aspects of their lives. Your contribution and data will help us better understand relations between human psychology and online behaviour." That is not quite the real purpose, is it?

**Aleksandr Kogan:** It is one of the purposes. Looking back, I should have been much more critical of the document. I did not write it—it was given to me—but where I failed is that I was not critical enough of reviewing that clause especially. I was assured that that was what we needed to do make it commercial.

**Paul Farrelly:** The primary purpose was political.

**Aleksandr Kogan:** That was one of the purposes. The way I understood it at that point is you write these things broadly so that you have broad scope to do what you will with the data. In truth, a variety of things were planned: one of them was political, another was academic and the third became some of the other commercial uses. If we had to do it again, I would have insisted to Mr Wylie that we add politics as a use case in that document.

Q1961 **Paul Farrelly:** It wasn't Christopher Wylie's company; it was yours and



## HOUSE OF COMMONS

Mr Chancellor's. Point 3 on the thisisyourdigitallife app is pretty much the same: the purpose of the application. But it is misleading—it is a misrepresentation.

**Aleksandr Kogan:** I think it is broad and not specific enough. You are asking why we did not outline specific-use cases, because politics is a specific-use case. I would argue that politics does fall under that, but it is a specific-use case, and I think we should have it.

Q1962 **Paul Farrelly:** I do not imagine that anyone who paid a couple of dollars would read this, but that is up to them. Lower down, in longer, denser paragraphs, you make it clear in both cases that whatever that primary purpose is, you can sell this data for any purpose whatsoever. How does that sit in terms of representing to people what you are doing?

**Aleksandr Kogan:** In which way, sir?

Q1963 **Paul Farrelly:** Prominently, telling the truth—what the real purpose behind the survey was.

**Aleksandr Kogan:** In terms of speaking the truth, the reality is, as you have pointed out, that very few people—if any—read this, just like very few if any people read terms of service. That is a major flaw we have right now: people just do not read these things. These things are written this way. Fundamentally, I made a mistake by not being critical about this and trusting the advice of another company. As you pointed out, GSR is my company and I should have got better advice and better guidance on what is and is not appropriate.

Q1964 **Paul Farrelly:** But immediately, the data that you were able to get from people taking your services was passed on to SCL Elections for a purpose that was not made clear in the application's terms and conditions—a purpose that never mentioned politics at all. That was completely misleading, if anyone had bothered to read it.

**Aleksandr Kogan:** Quite frankly, my understanding was that this was business as usual and normal practice for companies to write broad terms of service that did not provide specific-use cases. I will give you an example. I doubt Facebook's user policy says that users can be advertised for political purpose, but that it has broad language, to provide for any use cases they want. I agree with you that this does not seem right. Those changes need to be made.

Q1965 **Paul Farrelly:** Winding up, I want to come to your "60 Minutes" interview on 22 April. From the text of that, do you accept that you broke Facebook's terms and conditions?

**Aleksandr Kogan:** That was cut in an interesting way.

Q1966 **Paul Farrelly:** Is that a yes or a no?

**Aleksandr Kogan:** Did I break the terms of service?

**Paul Farrelly:** You accept that you broke them.

**Aleksandr Kogan:** I do not.

**Paul Farrelly:** You do not.

**Aleksandr Kogan:** I don't think they have a developer policy that is valid.

**Paul Farrelly:** That is a different answer.

**Aleksandr Kogan:** For you to break a policy it has to exist and really be their policy. The reality is that Facebook's policy is unlikely to be their policy.

Q1967 **Paul Farrelly:** It is about what it is in black and white. Do you accept that you broke the terms and conditions of Facebook, as laid out in black and white?

**Aleksandr Kogan:** I do not.

**Paul Farrelly:** Irrespective of how they enforced or whether or not—

**Aleksandr Kogan:** I just don't believe that is their policy. I mean, if somebody has a document that is not their policy, you cannot break something that is not really your policy. I would agree my actions were inconsistent with the language of this document. That is slightly different from what I think you are asking.

Q1968 **Paul Farrelly:** You should be a professor of semantics—Dr Kogan, is it?

**Aleksandr Kogan:** Yes.

Q1969 **Paul Farrelly:** I take it you accept that you have broken, in black and white, Facebook's terms and conditions—

**Aleksandr Kogan:** I do not, but continue.

Q1970 **Paul Farrelly:** How do you square that with the way you were operating? You are an academic at the University of Cambridge. The company was registered at Magdalene College. Did the University of Cambridge ever realise that your commercial venture, registered at one of their colleges, was breaking the terms and conditions of Facebook, one of the companies the university was dealing with?

**Aleksandr Kogan:** I suspect not, but—

Q1971 **Paul Farrelly:** You didn't tell them?

**Aleksandr Kogan:** That I was breaking the terms of service?

Q1972 **Paul Farrelly:** You didn't tell them?

**Aleksandr Kogan:** No, because I was not even aware of it. But, sir, on this issue, do you use Facebook?

Q1973 **Paul Farrelly:** I do.

**Aleksandr Kogan:** When you signed up, did you read their user policy?

Q1974 **Paul Farrelly:** No.

**Aleksandr Kogan:** Did you realise there was a user policy?

Q1975 **Paul Farrelly:** I didn't, but you have set up companies that have their own terms and conditions, so presumably you would know where you were with Facebook.

**Aleksandr Kogan:** I am just working through the experience. I think it is relevant to understand.

Q1976 **Paul Farrelly:** Let me ask my final question, another loose end: what was the purpose of Philometrics?

**Aleksandr Kogan:** We basically built a company where we had survey software. It is a bit like SurveyMonkey, if you have used that. The idea was to provide easy tools. The special thing, which is a little unusual, was that we were trying to figure out ways to forecast surveys. What we found through our experiences was that trying to predict Paul Farrelly is impossible—that is a waste of time—but there is value to trying to predict people's responses in aggregate. Philometrics was built, in a way, to try to help researchers and companies to predict people in aggregate, so they could better understand the diversity across people.

Q1977 **Paul Farrelly:** It was not related to the work you were doing for SCL Elections.

**Aleksandr Kogan:** No.

Q1978 **Simon Hart:** More on the Facebook relationship. I was reading the comments by Christopher Wylie when he gave evidence to us. He said, "I remember when—and I think this was around July 2014—Kogan was delayed for a couple of days because Facebook had throttled the app, so that it could not pull as much data. There was some problem with pulling as much data", etc. Is that true?

**Aleksandr Kogan:** I don't think it is. I have looked through all my records to try to find any mention of this, because when I saw that, I was very surprised. The only time I see any mention of data disruption is when Mr Wylie messages me in late June and says, "Hey, has the Facebook blackout affected us?" My response is: "What blackout? Wasn't even aware of it." I know reporters have asked Facebook about this issue, and Facebook has told the reporters this never happened.

Q1979 **Simon Hart:** So as far as you were concerned it never happened and there was no moment when you spoke to Facebook engineers—

**Aleksandr Kogan:** I don't know any engineers at Facebook.

Q1980 **Simon Hart:** So there was no dialogue between you and the company at the time. Why would Christopher Wylie invent that story?

**Aleksandr Kogan:** Mr Wylie has invented many things. You will have to ask him why he has done that.

Q1981 **Simon Hart:** We might do. I just wondered whether you would hazard a guess as to why he might have done it.

**Aleksandr Kogan:** I prefer not to speculate on his internal motives for his actions.

Q1982 **Simon Hart:** Fair enough. Was there ever a moment, leading up to the eventual divorce between you and Facebook, when Facebook engineers began to wonder? At what stage did they start to rumble you? At what stage did they think that you had started to breach the terms and conditions? How did the dialogue go?

**Aleksandr Kogan:** Going back to Mr Farrelly's questioning about the terms of service for Facebook as a developer, I believe I became aware that there was this inconsistency between their document and what we did in March 2015. Up to that point, I do not believe I was even aware of or had looked at the developer policy. I know that seems shocking and surprising, and this is what I was trying to get at before: the experience of a developer in Facebook is very much like the experience of a user on Facebook. When you sign up, there is this small print, easy to miss. When I made my app initially I was just an academic researcher. No company was involved yet. Then, when we commercialised it—so we changed the app—it was just something I completely missed. I do not have any legal resources. I relied on SCL. That was my mistake. The truth is that I relied on SCL to provide me this guidance on what was appropriate. By March 2015, we had begun to suspect that Mr Wylie may not be the most reputable person in the world and that we should question some of the advice he gave us.

Q1983 **Simon Hart:** What triggered that? Why did you reach that conclusion?

**Aleksandr Kogan:** We reached that conclusion because it was clear that he was duplicitous in his arrangement with SCL and also with us. If you recall, he entered into this contractual relationship with my company to exchange data sets. We honoured the agreement and he did not. It became clear as we looked at it that there might be a scheme there. We reached out to an IP lawyer at this point and got some guidance on the issue. In March 2015, we realised, "Hey, there is this inconsistency." Why I think this is still not Facebook's policy is that we were advised that Facebook's terms for users and developers are inconsistent and that it is not actually a defensible position for Facebook that this is their policy.

Q1984 **Simon Hart:** But they must have been checking your work.

**Aleksandr Kogan:** This is the remarkable thing about the experience of an app developer on Facebook. You can change the name, you can change the description, you can change the terms of service, and you just save changes. There is no obvious review process. We had a terms of service up on the Facebook platform—linked to the Facebook platform—that said we could transfer and sell data for at least a year and a half, and nothing was ever mentioned. It was only in the wake of the *Guardian* article that they came knocking.

Q1985 **Simon Hart:** When they did come knocking, wouldn't the easiest thing to have been to simply adjust what you were doing so that you did comply, or had that opportunity gone by then?

**Aleksandr Kogan:** There were two things I was trying to accomplish. They were still somebody I considered an ally and a friend. We collaborated with them. I was interested in trying to avoid any negative consequences for my students—as far as the publications were concerned. That was being held over me, that “We are going to pause all of the data sets that aren’t even related to the GSR project until this is resolved.” So I had a strong interest to just comply with them in any way I could, which we did. Their approach was, “Let’s ban the app.” I believe they might have even deleted their own records. That is the impression that I have got from talking to reporters about it.

Q1986 **Simon Hart:** I understand. My final question on this theme is once this had all happened, what measures did Facebook take to check that any material that they considered you had but should not have had was deleted?

**Aleksandr Kogan:** None.

**Simon Hart:** None at all.

**Aleksandr Kogan:** None at all.

Q1987 **Jo Stevens:** Dr Kogan, you mentioned earlier about some of the £230,000 that you have earned being spent on lawyers negotiating with Facebook. What were they negotiating?

**Aleksandr Kogan:** This is a difficult question, because I am under NDA, not to disclose the details of that agreement. But we are basically trying to protect ourselves and negotiate to make sure it was a happy resolution for everybody.

Q1988 **Jo Stevens:** Why have you got a non-disclosure agreement with Facebook?

**Aleksandr Kogan:** You will have to ask Facebook.

Q1989 **Jo Stevens:** Okay, so sticking with Facebook, you also said earlier that when you were doing work with data sets from Facebook in 2013-14, there was no requirement given to you, when you were given the data sets in the beginning, to delete them and you did not sign any agreement with them when the data was actually handed over to you. Can you explain the sequence of events from the point when Facebook first contacted you to delete the data that you had and what happened afterwards?

**Aleksandr Kogan:** I will speak a bit more broadly, because of the NDA. There was a request, as Facebook has already made publicly known, that they asked us to delete the data and certify that we deleted the data, and we went through that process where we deleted the data as best as we could—checked everywhere we could—and then certified that we had deleted that data.

Q1990 **Jo Stevens:** Can you remember when that was done?

**Aleksandr Kogan:** That was done during the first half of 2016.

Q1991 **Jo Stevens:** Did you also delete models and algorithms that have been derived from that data?

**Aleksandr Kogan:** Yes.

Q1992 **Jo Stevens:** So absolutely everything was deleted.

**Aleksandr Kogan:** As best as we could. We are now doing a review just to make sure that we did not miss anything in terms of remnants. But we haven't used any since then.

Q1993 **Jo Stevens:** What could you have missed?

**Aleksandr Kogan:** The things we could have missed are maybe a summary file somewhere, such as the average across these users about something, or maybe an email I missed that a dataset might have been in. We are trying to figure this out right now, just to make sure.

Q1994 **Jo Stevens:** Have you done that on your own initiative or at the request of Facebook?

**Aleksandr Kogan:** My own initiative.

Q1995 **Jo Stevens:** What further measures did Facebook take after they asked you to delete the data to ensure that you had deleted it? Was there any exchange of correspondence?

**Aleksandr Kogan:** Once we certified, the matter was generally closed. As we have talked about before, it is actually a legitimately difficult thing to achieve because any sort of audit will never catch bad actors. You cannot stop somebody putting the data on a hard drive and sticking it under their mattress. I think Facebook is aware of that. It is a tech company.

Q1996 **Jo Stevens:** Your company did not put the data on a hard drive and stick it under a mattress.

**Aleksandr Kogan:** We didn't. Look, this has been a very painful experience because when I entered into all of this, Facebook was a close ally and I was thinking this would be helpful to my academic career and my relationship with Facebook. It has very clearly done the complete opposite. I had no interest in becoming an enemy or in being antagonised by one of the biggest companies in the world that could, even if it is frivolous, sue me into oblivion. So we acted entirely as they requested.

Q1997 **Jo Stevens:** Right, okay. Just going back to this NDA, I know you say you cannot talk about the details of it, but was it only you that was required to sign an NDA or were any of your co-directors required to?

**Aleksandr Kogan:** I can't talk about that. Ask Facebook. I understand that is frustrating.

Q1998 **Jo Stevens:** Drawing on one of the questions earlier, it seems very odd to me that you have been personally attacked by Facebook and criticised very severely and yet your co-directors have not and you were all doing exactly the same work.

**Aleksandr Kogan:** I think it is odd from a fairness perspective. It is honestly just not that odd from a PR perspective. That is the reality of it.

Q1999 **Jo Stevens:** Can you tell us when the NDA was signed?

**Aleksandr Kogan:** I cannot.

**Paul Farrelly:** It is like a super-injunction.

**Aleksandr Kogan:** I know there is a concept of parliamentary privilege, but we just don't think it extends over to the United States or, at least, there is an ambiguity.

Q2000 **Jo Stevens:** So the NDA was signed in the United States.

**Aleksandr Kogan:** I cannot talk about that agreement.

**Jo Stevens:** Okay.

**Aleksandr Kogan:** I am sorry. I want to.

Q2001 **Jo Stevens:** What contact did you have with SCL or Cambridge Analytica to recover the data that you had passed to them?

**Aleksandr Kogan:** As you can imagine, I have no ability to enforce anything. We requested that they delete all the data.

Q2002 **Jo Stevens:** When did you do that?

**Aleksandr Kogan:** Again, during 2016. They confirmed to me orally that they were going to delete the data, but beyond that—

Q2003 **Jo Stevens:** And that was the end of the dialogue.

**Aleksandr Kogan:** Yes. In my understanding, Facebook spoke to them and Facebook did get certification from them that they deleted the data—that is public knowledge. Beyond that, it is very difficult to make sure that SCL did not put the hard drive under the mattress, so to speak.

Q2004 **Jo Stevens:** Can you confirm to us for the record that you don't continue to hold any data or derivatives from your Facebook apps?

**Aleksandr Kogan:** I can confirm that as far as we know, we did the best we could to delete it, and we are just double-checking. I do not want to say definitively that we don't have it until I am done checking. We are obviously in a tricky spot where I cannot delete anything right now given the nature of the situation, but I can confirm for the record that I have not used anything from this data and I knowingly did not keep anything from this data following the request, and I am just making sure we are thorough right now.

Q2005 **Jo Stevens:** 100%. Okay. Finally, for the record, how many other companies did you sell or give your Facebook data to?

**Aleksandr Kogan:** The data was given to the following entities. It was given to SCL. The dataset that was given to SCL was restricted in the sense that there was very little raw Facebook data. The data was given to



## HOUSE OF COMMONS

Eunoia, Chris Wylie's company, and that was more broad, because we gave them the raw Facebook data, so that is probably the biggest dataset. We gave an anonymous version of the dataset to the University of Cambridge through a data transfer agreement. Then we shared a derivative dataset with a researcher at the University of Toronto at the University of Toronto who I was collaborating with, but this dataset was based only on the people who had taken the survey, not their friends, and it was all predicted data rather than any sort of raw Facebook data, so it wasn't the actual profiles, and it was, again, all fully anonymised.

I can confirm that we did our absolute best to delete the dataset from the University of Cambridge, because that was my lab and I could control that, and I can confirm that the researcher that we gave the dataset to also deleted it.

Q2006 **Jo Stevens:** But not SCL, Eunoia and the University of Toronto. You can't—

**Aleksandr Kogan:** No, also Toronto I can confirm. I am well assured that that happened. The two that I am less confident in are SCL and Eunoia.

Q2007 **Jo Stevens:** When you share that data, do you have written agreements with those organisations?

**Aleksandr Kogan:** Yes.

Q2008 **Jo Stevens:** When you enter into that arrangement?

**Aleksandr Kogan:** Yes, absolutely.

Q2009 **Jo Stevens:** And do your written agreements contain any clauses relating to the necessity to delete data at the end of the project?

**Aleksandr Kogan:** At the end of projects, no. That was unusual, I would say, because typically we would keep a dataset because there could be further usage for it in other areas.

Q2010 **Jo Stevens:** You were pretty upset that Facebook wanted you to delete the data, weren't you? You wanted to be able to keep it for your PhD projects and research.

**Aleksandr Kogan:** Yes.

Q2011 **Jo Stevens:** So in future do you think you would include deletion clauses in agreements?

**Aleksandr Kogan:** For others?

**Jo Stevens:** Yes, if you are handing over data to other companies.

**Aleksandr Kogan:** I doubt I'm going to be in this position in the future, to be honest. I think in the future we would be very careful before we gave data to anybody and we would have very strong restrictions on how the data was given and limits on scope. I would have to think about it, in terms of deletion. That's just not an issue I have really considered in

terms of timeboxing it, but I think it would be very likely it would be something I would want to include.

**Jo Stevens:** Thank you, Dr Kogan.

Q2012 **Chair:** Could you tell us a little bit about your work with AggregateIQ as part of your work with SCL?

**Aleksandr Kogan:** There is nothing to talk about. I had never even heard of them up until about a month ago.

Q2013 **Chair:** Because there was the project for the Bolton Super PAC AggregateIQ were working on, with SCL as well.

**Aleksandr Kogan:** This is something that I know now. That was not something I was ever aware of.

Q2014 **Chair:** Did you ever work with the Ripon tool that AggregateIQ developed? When you were processing the data that you were gathering from the surveys, how was that being used? You said earlier that there were tools that had been developed at the University of Cambridge that could have been used to process the data but weren't. We know that Ripon was the tool that was used by SCL—created for them by AggregateIQ to process data. So were you, if you like, loading the data that you were gathering into Ripon for them to use for the Bolton Super PAC?

**Aleksandr Kogan:** I don't really know what Ripon does. I can tell you what we did. We used a piece of software called R—just statistical software. We built models; we made predictions; and we saved them into basically Excel files. There were CC files, but I think it was Excel files. We would zip this up, password-protect it, and send it off to SCL, and separately I would provide them with the password. What they did with these Excel files, I have no idea.

Q2015 **Chair:** Who was kind of your handler at SCL? Who was the person that you interacted with?

**Aleksandr Kogan:** There was a woman named Sabhita, who was the project manager, who I interacted with mostly.

Q2016 **Chair:** Indeed, she is the author of the email that I quoted from earlier, so obviously she is referring directly to conversations she has had with you. But other than that? Obviously, Chris Wylie you have discussed, but were there other members of the team you met with regularly?

**Aleksandr Kogan:** I met a couple of times with a few other folks from the data science team—again, for their privacy, I prefer not to bring their names up, but there were a few of the data scientist folks for sure. Sabhita is a project manager, but the other folks are actually handling the data, so I would be interacting with them mostly.

Q2017 **Chair:** Did you work with Jeff Silvester at all?

**Aleksandr Kogan:** That name does not ring a bell.

Q2018 **Chair:** I just wonder whether there are people that we now know as AggregateIQ or AIQ but might at the time just have been considered to be part of SCL.

**Aleksandr Kogan:** There's nobody I'm aware of from AggregateIQ that was part of SCL that rings a bell in any way. I think I can account for pretty much everybody I interacted with and they are not in any way involved, as far as I know.

Q2019 **Chair:** Were the SCL members you worked with based in the UK or the US?

**Aleksandr Kogan:** UK.

Q2020 **Chair:** They are all UK.

**Aleksandr Kogan:** Yeah. I mean, almost everybody I interacted with was a former PhD student from Cambridge.

Q2021 **Chair:** For the record, it would be quite useful if you could explain to us why Facebook data was so important to the work you were doing.

**Aleksandr Kogan:** My work as an academic, or my work for SCL?

Q2022 **Chair:** Both, if they are different.

**Aleksandr Kogan:** As a social psychologist I am interested in understanding people. One big problem we have had in the field is that we typically recruit undergraduates in psychology departments, and study them and try to understand people. As you can imagine, an undergrad in a psych programme is not the most representative human being alive.

**Chair:** Thank god.

**Aleksandr Kogan:** Yes—it's a pretty big flaw. We do this because it is expensive to do anything else. In the field there has been long disillusion with this, but it is a big problem. The interest in social media and big data in general for me was: can we overcome this fundamental problem? Can we get to studying people? Facebook was interesting for two reasons. First, it had a relatively open API, so you could get a lot of data on a lot of people. Secondly, the Facebook user base is very representative. I think in the United States it is 70% or 80% of the adult population. Few other data sets are that representative of people. That is why I was personally interested in this, from an academic angle. For SCL, the interest in Facebook was that it could gather a lot of data because it is a pretty open API. Beyond that there is no real reason it has to be Facebook.

Perhaps I can add a bit more, because I think a little more context matters. We already talked about how it was interested in acquiring the myPersonality project. Another thing happening at the same time was that it was thinking of going at a measure of personality through phone surveys. The plan, I believe, was to call people up and ask them 10 items to measure their personality, which is just very noisy. This was an alternative where you could use online methods but still get a pretty big



## HOUSE OF COMMONS

sample, and you can balance out the randomness—you can do that with phones, but you have a much longer survey.

**Q2023 Chair:** A number of pieces of academic research have been conducted looking at the predictive qualities of Facebook data. Do you believe it is superior to other commonly available data sets?

**Aleksandr Kogan:** A lot of the ideas about accuracy are grossly overstated. Can I walk you through this because it is such an important point? I have prepared some slides for you. It might be unusual to do a stats lecture, but it is short.

**Q2024 Chair:** Given that we are slightly—we may not have time to do that now, but we could accept the slides you have given us as a written submission to the Committee.

**Aleksandr Kogan:** Sure. Let me just high-level for you. Imagine you were trying to predict people's age. America is a great place because we know the numbers—the average is 41 and we have the spread. If you did a random guess, you would be off by about 27 years, which is pretty bad. If you guessed that everybody in America is 41 years old, you would be off by about 19 years. If you used Facebook data and had the same accuracy that we have for our personality data, you would be off by 18 years, but the problem would be that it would tell you that most people are middle-aged. When you correct for that, you would get a situation where you think that most people are—you would be wrong by about 22 years. The idea that this data is accurate is scientifically ridiculous. The idea that, even if you had a lot more data, you can make it super-accurate, is also pretty silly once you work through—even if you had very high correlations—what it means for actual accuracy.

The project, quite frankly, if the goal is micro-targeting using Facebook ads, makes no sense. It is not what you would do. If you want to do a project where you micro-target people using Facebook ads, you use the Facebook ad platform and target 100% of the population rather than 15%. We haven't talked about numbers, but we gave SCL 30 million people. Why would you want to target only 15% of the population and use only their page likes to do it, when you could target everybody using much more information? You just don't need this data to do that; the Facebook platform gives you every ability to do that, even if you are interested in psychographics.

**Q2025 Chair:** So in your view you could just micro-target through Facebook as a platform, rather than developing your own database.

**Aleksandr Kogan:** Yes—perhaps you will allow me to explain how. The proper way to do this project if you wanted to do it right now—suppose you wanted to say, "I want to target people with ads related to personality; I want to target really extroverted people". All you do is recruit 10,000 people. You don't collect any Facebook data at all. You ask them for their email addresses and you ask them to fill out a personality survey—10,000 people. Then you say, "Who are the 2,000 who are the most extroverted?" Then you take their email addresses and you go on the



## HOUSE OF COMMONS

Facebook app platform and you say, "Facebook, please build me a lookalike audience." What they are going to do is take those 2,000 email addresses, find the people on Facebook who match those 2,000 email addresses, and then use all the data that they have—or a lot of the data they have—to figure out who all the people like this are. That way you will find everybody who is an extrovert, potentially, based on the other predictions, and you would use a lot more of the Facebook data, rather than the page likes. My understanding is that they try to run Facebook ads from this. This project is just a waste of time and a colossal waste of money for them from that perspective.

**Q2026 Chair:** But isn't the point the linking of someone's Facebook data to other personality indicators as well, and the ability to then go back to Facebook? We took evidence on this last week—people saying that actually once you have your working dataset of individuals, you can then use Facebook tools to identify more people who are like the group in that set.

**Aleksandr Kogan:** You don't even need to do that. Like you said, just gather email addresses for 2,000 people and get who the real extroverts are. The problem is that the predictions we made about personality are incredibly inaccurate—really noisy. If you try to take these highly noisy predictions, and select who the extroverts are, you are going to select a lot of people who are not extroverts. When you go back to Facebook and say, "Hey, please build me a lookalike audience on this really noisy group," it will give you a really noisy group. It just does not make sense.

**Q2027 Chair:** In that case, what was the value of the project that you were doing for SCL?

**Aleksandr Kogan:** Given what we know now, nothing. Literally, nothing.

**Q2028 Chair:** They hired you as an expert, and you designed a survey that you say is worthless. Is that correct?

**Aleksandr Kogan:** It is all about the use case. I was very surprised to learn that what they wanted to do was run Facebook ads. This was not mentioned back then; they just wanted a way to measure personality for many people. If the use case you have is Facebook ads, it is just incompetent to do it this way.

**Q2029 Chair:** Because your view is you can just do it directly through Facebook?

**Aleksandr Kogan:** Yes. Taking this dataset, you will be able to target 15% of the population and use a very small segment of the Facebook data, which is page likes, to try to build personality models. Why do that when you could very easily target 100% and use much more of the data? It just does not make sense.

**Q2030 Chair:** I was interested in how you sold this to them. Do you look back at your work then and say that you were mistaken in the value you placed on this type of work?



**Aleksandr Kogan:** The idea that I “sold them” is a strong statement. They were already really interested in personality. I told you that they were going to call up people and try to measure their personality through the phone survey. What we gave them was an alternative to tackle this problem.

What they did with it is the surprise, because if they turned around and tried to run Facebook ads on the back of this, that is where the incompetence comes in—but that is not something they told me. Quite frankly, I did not know a lot on Facebook ads back then. I was not an expert on how the ad platform works. Now I know much more; I have had more experience with it. I might have missed that, but I would hope, if this is their plan, that somebody there looks at how to run Facebook ads and they would figure this out.

Q2031 **Chair:** But whose idea was it to have people give their Facebook data as part of completing the survey?

**Aleksandr Kogan:** That was myself and the two psychometrics members.

Q2032 **Chair:** Why did you recommend that?

**Aleksandr Kogan:** It was just an alternative to going out and phoning people up. The phoning people up idea is pretty difficult because you are asking very few questions to measure a complex thing. This was a more efficient way to do it online.

Q2033 **Chair:** Yes, but obviously there was a lot more benefit to it than just identifying people, or just reaching people through an easy mechanism like Facebook, wasn't there? You knew full well that by doing it in this way you would gather a lot more data.

**Aleksandr Kogan:** Yes, it is exactly that. It was a more efficient way of gathering a big dataset.

Q2034 **Chair:** Because you were gathering the dataset not just for people who took the survey, but all their friends as well.

**Aleksandr Kogan:** Exactly.

Q2035 **Chair:** And you are saying that there was no particular value to that?

**Aleksandr Kogan:** If the goal is for you to run Facebook ads, the friends data is useless.

Q2036 **Chair:** What is the value of it then?

**Aleksandr Kogan:** Of this dataset? I think it is actually in the aggregate form. I will give you an example of what I was doing with it. After we did the dataset, we were working on questions like: “Are people who are kinder happier?” We found that it really depends. In states like New York, kinder people are less happy on average. In states like Utah, people who are kinder tend to be happier.

Then if you look at the friends information—we predict people's kindness from their friends—you find that it is about your social network. Folks who



## HOUSE OF COMMONS

are kinder, and are surrounded by kinder people, tend to be happier, but if you are surrounded by not kind people, it is worse. It is questions like that, which have nothing to do with running ads; they are very basic science questions about human nature.

**Q2037 Chair:** You said at the beginning that you were getting aggregated data from Facebook anyway, so what was the point of creating your own app to get the dataset?

**Aleksandr Kogan:** You cannot do this. To do this, you have to look at the relationship at the individual level. You need to know, for example, 10,000 people in New York and their happiness and kindness scores. You need individual-level data to do this.

**Q2038 Chair:** You have given your view on it. I think others would ascribe more value to the data that you were gathering and the importance of gathering it in the way that you were. You stated your position on that.

I just want to ask a couple of questions about Mark Zuckerberg's initial statement on his Facebook page on 21 March. He, in a typically carefully worded statement, walks us through the chronology of events of your work as he saw it. I just wanted to ask about that.

He says that in 2014 they made changes to the way the platform works "to prevent abusive apps", as he called it. If by abusive apps he means apps that gather Facebook friends' data, that was completely within the terms and conditions of the site at the time, so I do not know how he can describe that as being abusive.

**Aleksandr Kogan:** I would agree with that.

**Q2039 Chair:** He also said that, as a result of the changes: "We also required developers to get approval from us before they could request any sensitive data from people." Presumably that means that, before 2014, you didn't need their approval to gather sensitive data about Facebook users?

**Aleksandr Kogan:** Yes. You would just say what you wanted and they would give it to you.

**Q2040 Chair:** He said: "It is against our policies for developers to share data without people's consent". You disputed that earlier on. Did you believe that, because you had your own terms and conditions for your app, you therefore had asked their consent? Was that your view?

**Aleksandr Kogan:** That was certainly my feeling, or my understanding, at the time. I would have a different view now, but that was my perspective at the time.

**Q2041 Chair:** When Facebook brought these changes in, in 2014, did you believe that apps that had been developed and launched before those changes could still work to the old rules?

**Aleksandr Kogan:** That was a fact. Old apps were given a year to continue to operate under the old rules. Facebook basically said that there

were two versions of the API: the original and the new version. The old version, which ran for another year, was allowed to operate as before.

**Q2042 Chair:** So your apps were incredibly valuable for people who wanted to gather Facebook data at that time, weren't they? They could still enjoy the grandfathered rights of the old APIs, rather than the new ones. Did people speak to you, or did you speak to other organisations, about trying to make the most of the tools you had developed before that time ran out?

**Aleksandr Kogan:** It was literally only the SCL project that we ever did with anybody, in terms of commercial activities. We didn't talk to anybody else afterwards or collect more data. That was it.

**Q2043 Chair:** I want to ask about some of your other work as well. While you were at the University of Cambridge, you also worked at the University of St Petersburg, didn't you?

**Aleksandr Kogan:** Yes, I had a loose affiliation there.

**Q2044 Chair:** In what time period did you do that work?

**Aleksandr Kogan:** In, I believe, the summer of 2013, a couple of my friends wanted to visit Saint Petersburg. I went along, because Russia is actually a pretty hard place to go without a Russian speaker, and I can more or less speak Russian. While there, I visited the university to say hi. That is kind of a typical thing I do in most places I visit.

After that, they said we should set up some kind of collaboration. They invited me out to do some talks—a year later, I think. At some point, some of the researchers there applied for a grant from the university, and I believe they put my name on it because they thought it would help. They got it. I didn't read or write the application. I can't write in Russian. I can barely read in Russian—very slowly.

They got that grant. I was then invited out to do a couple more talks. I did a workshop on statistics. That was it. I think we had maybe one or two meetings where I gave them a bit of advice about their project. I was never part of the data collection and I never had access to the data in any way.

**Q2045 Chair:** You described some of the work that was done. What was the thesis that you were working on? What was the subject matter?

**Aleksandr Kogan:** I won't say me, because it was really them; I don't want to take credit for somebody else's work. I believe they were tackling how to curb cyber-bullying and how to stop people being mean to each other online.

**Q2046 Chair:** When was that work being done?

**Aleksandr Kogan:** I want to say it was between 2014 and 2016. I think that's right, but I would have to double check.

**Q2047 Chair:** So following on from the work you were doing with SCL?

**Aleksandr Kogan:** To an extent, yes.

**Chair:** Overlapping a little bit.

**Aleksandr Kogan:** For sure.

Q2048 **Chair:** Was Cambridge University aware of the work you were doing in St Petersburg?

**Aleksandr Kogan:** Of course. Before I accepted any position, I went to my department head at the university to make sure that everything was fine.

Q2049 **Chair:** Are there any requirements to notify other authorities that you are conducting research work in Russia?

**Aleksandr Kogan:** Not that I am aware of.

**Chair:** So you certainly did not. So you only informed—

**Aleksandr Kogan:** The university.

Q2050 **Chair:** Who else in Russia, outside the university, was aware of the work you were doing?

**Aleksandr Kogan:** I have no idea. I only interacted with three or four researchers there. Beyond that, I have no idea.

Q2051 **Chair:** Russia has been the subject of our investigation as well, and there continue to be questions around that. Is it possible at all that, because you travelled with a laptop that contained data and information, people in Russia could have gained access to or benefited from the work you were doing for SCL?

**Aleksandr Kogan:** You don't travel with the dataset on your laptop. It is too much data. It lived in a server in Portland. That just doesn't make sense.

Q2052 **Chair:** There is no way for that to be remotely accessed?

**Aleksandr Kogan:** If somebody wants to hack Amazon, go for it, but then just hack Facebook. I think that is mostly a ridiculous idea.

**Chair:** Okay. Thank you.

Q2053 **Paul Farrelly:** Your accent is American, and then we have Russia coming up. You moved to the States as a child.

**Aleksandr Kogan:** As a seven-year-old.

**Paul Farrelly:** From?

**Aleksandr Kogan:** I was born in Moldova. And then—

**Paul Farrelly:** A place I am still trying to get to. It has some great wine.

**Aleksandr Kogan:** I haven't been for a long time. I actually don't know much about it. We emigrated when I was seven years old from Moldova.

Q2054 **Paul Farrelly:** Out of curiosity, what do your mother and father do?

**Aleksandr Kogan:** My father was a programmer and my mother is a seamstress/fashion designer. The motives for our move were that the country had just fallen apart and we were getting death threats on account of my father being Jewish. We were either going to move to Israel or to the United States. I believe we went through as Jewish refugees. If my recollection is correct, one of the Jewish programmes in the United States sponsored our plane tickets.

Q2055 **Paul Farrelly:** Has your work in St Petersburg now finished?

**Aleksandr Kogan:** As far as I know.

Q2056 **Paul Farrelly:** Which means that if St Petersburg comes up, we have to switch the alarm bells off, because it is just pure coincidence.

**Aleksandr Kogan:** It is pure coincidence.

Q2057 **Jo Stevens:** Do you know whether any GSR data was used by the Vote Leave campaign in the referendum?

**Aleksandr Kogan:** I don't know, but I don't know why they would. We only gave them data for people who reported their location in the United States.

Q2058 **Jo Stevens:** Your team is non-US, isn't it?

**Aleksandr Kogan:** What do you mean by non-US?

**Jo Stevens:** Are your team all based in the United States?

**Aleksandr Kogan:** The GSR team? No, we were in Cambridge.

Q2059 **Jo Stevens:** You're over here, but you're processing US election data, aren't you?

**Aleksandr Kogan:** We're processing Facebook data.

Q2060 **Jo Stevens:** Outside the jurisdiction of the US—so US citizens' data being processed outside the US.

**Aleksandr Kogan:** This bit is a little tricky. The servers, where they were set up, were in Portland, Oregon. Keep in mind that there is now the cloud. The database was in Portland, Oregon. We collected the data through the servers in Portland, Oregon. I am not exactly sure where the modelling was done, either here or there.

Q2061 **Jo Stevens:** So it was your company doing the modelling, but you do not know where it was done.

**Aleksandr Kogan:** It was either here or in the US. It is entirely possible that it was here. I think it would probably be safer—when we delivered the files, especially the second version, we gave a physical hard drive to SCL.

Q2062 **Jo Stevens:** Do you accept that if it is processed here, it is processed outside the US jurisdiction—that US citizens' election data is processed outside the jurisdiction?

**Aleksandr Kogan:** The only thing is the elections bit. Data is certainly being processed outside the United States, but I am not sure that I would characterise it as elections data. There is no election result or anything like that. I just do not know the technical definition of election data.

Q2063 **Jo Stevens:** Okay—political.

**Aleksandr Kogan:** Even there, I don't—I fear that is a technical term in terms of how it is defined. I do not want to commit to legal language that means a specific thing, when we are speaking colloquially, if that makes sense.

Q2064 **Jo Stevens:** Okay. I will have one more try on the NDA. Was it signed pre or post the data breach in 2015?

**Aleksandr Kogan:** What do you mean by data breach?

Q2065 **Jo Stevens:** The Facebook data breach. The Facebook users data being passed and used through GSR, SCL, Cambridge Analytica?

**Aleksandr Kogan:** I cannot answer that question.

Q2066 **Jo Stevens:** I thought that might be the answer. Okay. It was worth a try.

**Aleksandr Kogan:** You can piece together the timeline.

Q2067 **Chair:** Just a few more questions, Dr Kogan. To cover it for the record, who funded the work that was done through the university of Saint Petersburg?

**Aleksandr Kogan:** The university.

Q2068 **Chair:** Often with universities you apply to a particular grant pot at a particular institution. Can you elaborate a little bit more about where the money came from within the university?

**Aleksandr Kogan:** I honestly don't know. I have read that the Russian Government gave a block grant to the university and then they divvy it up, but it's pretty normal practice in every country.

Q2069 **Chair:** So the Russian Government had given the funding to the university. Normally it is with the view of commissioning certain types of work in a subject that it has an interest in.

**Aleksandr Kogan:** In the UK and the US, that is how it works. In Russia I think it is much more generic. It is a big block grant and then the university is responsible for commissioning specific projects. That is a distinction between the two.

Q2070 **Chair:** But this was a piece of work on cyber bullying funded by the Russian Government, which might make you wonder what their particular interest in that was.

**Aleksandr Kogan:** Most work that is funded by—

Q2071 **Chair:** How to do it rather than how to stop it.

**Aleksandr Kogan:** You can make the same arguments about the UK and the US Governments funding anything. Both countries are very famous for their spies. That is a big leap.

Q2072 **Chair:** Obviously, a lot of work was done to analyse the interference of Russian agencies in foreign elections and the creation of networks of bot accounts and trolls who intimidate people.

**Aleksandr Kogan:** Of course, and there is a long history of the United States interfering in foreign elections and doing the exact same thing. Unless you want to argue that—

Q2073 **Chair:** Are you saying it is equivalent? Are you saying that the work of the Russian Government is equivalent to the work of the United States Government and you could not really distinguish between the two? Is that your opinion?

**Aleksandr Kogan:** In general I would say that of the Governments that are most high profile, I am dubious about the moral scruples of their activities through the long history of the UK, the United States and Russia. Trying to equate them is a bit of a silly process, but certainly all these countries have engaged in activities that are covert that people feel uncomfortable with. To try to link academic work that is basic science to that, if you are going to go down the Russian line, we have to go down the UK and US line in the same way, although I understand Russia is a hot button topic right now. Most people in Russia and the United Kingdom are not involved in spycraft. They are just living lives.

**Chair:** I am not aware of UK Government agencies that have been interfering in foreign elections—

**Aleksandr Kogan:** That does not mean it is not happening. They could just be better at it.

Q2074 **Chair:** If you have been funded by the Russian Government to look at cyber bullying, presumably a thesis was produced. Was that submitted to the Government as well as the university?

**Aleksandr Kogan:** I have no idea. I think they published some academic papers on it, but, since they are in Russian, I was not really involved.

Q2075 **Chair:** So you didn't see the end product.

**Aleksandr Kogan:** I think they might have sent one of the papers to me once. I didn't read it because I can't read Russian very well.

Q2076 **Chair:** What aspects of cyber bullying did it cover?

**Aleksandr Kogan:** I truly don't know the exact details of that project. I don't know the final results. The methodology I loosely understand and remember. Just keep in mind I was a name on a grant rather than an active participant and collaborator on this.

Q2077 **Chair:** How long did Dr Michal Kosinski work with you at GSR?

**Aleksandr Kogan:** I'm sorry?

Q2078 **Chair:** How long was Dr Michal Kosinski working with you at GSR?

**Aleksandr Kogan:** He never worked at GSR.

Q2079 **Chair:** He never did at all?

**Aleksandr Kogan:** No.

Q2080 **Chair:** Have you worked with him on any other projects?

**Aleksandr Kogan:** We were collaborating for a while in 2013 on academic work. I don't think it ever went anywhere. He was going to be part of the SCL project, as we have talked about before.

Q2081 **Chair:** But that ended up not happening.

**Aleksandr Kogan:** Exactly.

Q2082 **Chair:** What was the nature of the academic projects you collaborated on?

**Aleksandr Kogan:** Facebook data. We worked on various angles there. My lab had access to the myPersonality dataset that Dr Kosinski and Dr David Stillwell ran. So he had a variety of different project ideas. I don't think we ever landed on anything that was effective.

Q2083 **Chair:** I am sure that he posted on his Facebook page that he visited the Russian Prime Minister to brief him on some of the work being done by the psychometric centre at Cambridge. Were you aware of that?

**Aleksandr Kogan:** I was not.

Q2084 **Chair:** And you don't know what particular projects or work he discussed at that meeting?

**Aleksandr Kogan:** No idea.

Q2085 **Chair:** I presume he could have discussed some of the work he had done with you.

**Aleksandr Kogan:** I do not know why he would, given that it was so immaterial. When did he visit?

Q2086 **Chair:** I am sure I can get the dates. I don't have the printout in front of me. That probably concludes our questions, unless anyone else wants to come in. Dr Kogan, thank you very much for your evidence today. We are very grateful.