

The Data That Turned the World Upside Down

HG

HANNES GRASSEGER AND MIKAEL KROGERUS

Jan 28 2017, 3:15pm

Psychologist Michal Kosinski developed a method to analyze people in minute detail based on their Facebook activity. Did a similar tool help propel Donald Trump to victory? Two reporters from Zurich-based Das Magazin went data-gathering.

An earlier version of this story appeared in [Das Magazin](#) in December.

On November 9 at around 8.30 AM., Michal Kosinski woke up in the Hotel Sunnehüs in Zurich. The 34-year-old researcher had come to give a lecture at the Swiss Federal Institute of Technology (ETH) about the dangers of Big Data and the digital revolution. Kosinski gives regular lectures on this topic all over the world. He is a leading expert in psychometrics, a data-driven sub-branch of psychology. When he turned on the TV that morning, he saw that the bombshell had exploded: contrary to forecasts by all leading statisticians, Donald J. Trump had been elected president of the United States.

For a long time, Kosinski watched the Trump victory celebrations and the results coming in from each state. He had a hunch that the outcome of the election might have something to do with his research. Finally, he took a deep breath and turned off the TV.

On the same day, a then little-known British company based in London sent out a press release: "We are thrilled that our revolutionary approach to data-driven communication has played such an integral part in President-elect Trump's extraordinary win," Alexander James Ashburner Nix was quoted as saying. Nix is British, 41 years old, and CEO of Cambridge Analytica. He is always immaculately turned out in tailor-made suits and designer glasses, with his wavy blonde hair combed back from his forehead. His company wasn't just integral to Trump's online campaign, but to the UK's Brexit campaign as well.

Of these three players—reflective Kosinski, carefully groomed Nix and grinning Trump—one of them enabled the digital revolution, one of them executed it and one of them benefited from it.

How dangerous is big data?

Anyone who has not spent the last five years living on another planet will be familiar with the term Big Data. Big Data means, in essence, that everything we do, both on and offline, leaves digital traces. Every purchase we make with our cards, every search we type into Google, every movement we make when our mobile phone is in our pocket, every "like" is stored. Especially every "like." For a long time, it was not entirely clear what use this data could have—except, perhaps, that we might find ads for high blood pressure remedies just after we've Googled "reduce blood pressure."

On November 9, it became clear that maybe much more is possible. The company behind Trump's online campaign—the same company that had worked for Leave.EU in the very early stages of its "Brexit" campaign—was a Big Data company: Cambridge Analytica.

To understand the outcome of the election—and how political communication might work in the future—we need to begin with a strange incident at Cambridge University in 2014, at Kosinski's Psychometrics Center.

Psychometrics, sometimes also called psychographics, focuses on measuring psychological traits, such as personality. In the 1980s, two teams of psychologists developed a model that sought to assess human beings based on five personality traits, known as the "Big Five." These are: openness (how open you are to new experiences?), conscientiousness (how much of a perfectionist are you?), extroversion (how sociable are you?), agreeableness (how considerate and cooperative you are?) and neuroticism (are you easily upset?). Based on these dimensions—they are also known as OCEAN, an acronym for openness, conscientiousness, extroversion, agreeableness, neuroticism—we can make a relatively accurate assessment of the kind of person in front of us. This includes their needs and fears, and how they are likely to behave. The "Big Five" has become the standard technique of psychometrics. But for a long time, the problem with this approach was data collection, because it involved filling out a complicated, highly personal questionnaire. Then came the Internet. And Facebook. And Kosinski.

Michal Kosinski was a student in Warsaw when his life took a new direction in 2008. He was accepted by Cambridge University to do his PhD at the Psychometrics Centre, one of the oldest institutions of this kind worldwide. Kosinski joined fellow student David Stillwell (now a lecturer at Judge Business School at the University of Cambridge) about a year after Stillwell had launched a little Facebook application in the days when the platform had not yet become the behemoth it is today. Their MyPersonality app enabled users to fill out different psychometric questionnaires, including a handful of psychological questions from the Big Five personality questionnaire ("I panic easily," "I contradict others"). Based on the evaluation, users received a "personality profile"—individual Big Five values—and could opt-in to share their Facebook profile data with the researchers.

Followers of Lady Gaga were most probably extroverts, while those who "liked" philosophy tended to be introverts.

Kosinski had expected a few dozen college friends to fill in the questionnaire, but before long, hundreds, thousands, then millions of people had revealed their innermost convictions. Suddenly, the two doctoral candidates owned the largest dataset combining psychometric scores with Facebook profiles ever to be collected.

The approach that Kosinski and his colleagues developed over the next few years was actually quite simple. First, they provided test subjects with a questionnaire in the form of an online quiz. From their responses, the psychologists calculated the personal Big Five values of respondents. Kosinski's team then compared the results with all sorts of other online data from the subjects: what they "liked," shared or posted on Facebook, or what gender, age, place of residence they specified, for example. This enabled the researchers to connect the dots and make correlations.

Remarkably reliable deductions could be drawn from simple online actions. For example, men who "liked" the cosmetics brand MAC were slightly more likely to be gay; one of the best indicators for heterosexuality was "liking" Wu-Tang Clan. Followers of Lady Gaga were most probably extroverts, while those who "liked" philosophy tended to be introverts. While each piece of such information is too weak to produce a reliable prediction, when tens, hundreds, or thousands of individual data points are combined, the resulting predictions become really accurate.

Kosinski and his team tirelessly refined their models. In 2012, Kosinski proved that on the basis of an average of 68 Facebook "likes" by a user, it was possible to predict their skin color (with 95 percent accuracy), their sexual orientation (88 percent accuracy), and their affiliation to the Democratic or Republican party (85 percent). But it didn't stop there. Intelligence, religious affiliation, as well as alcohol, cigarette and drug use, could all be determined. From the data it was even possible to deduce whether someone's parents were divorced.

The strength of their modeling was illustrated by how well it could predict a subject's answers. Kosinski continued to work on the models incessantly: before long, he was able to evaluate a person better than the average work colleague, merely on the basis of ten Facebook "likes." Seventy "likes" were enough to outdo what a person's friends knew, 150 what their parents knew, and 300 "likes" what their partner knew. More "likes" could even surpass what a person thought they knew about themselves. On the day that Kosinski published these findings, he received two phone calls. The threat of a lawsuit and a job offer. Both from Facebook.

Only weeks later Facebook "likes" became private by default. Before that, the default setting was that anyone on the internet could see your "likes." But this was no obstacle to data collectors: while Kosinski always asked for the consent of Facebook users, many apps and online quizzes today require access to private data as a precondition for taking personality tests. (Anybody who wants to evaluate themselves based on their Facebook "likes" can do so [on Kosinski's website](#), and then compare their results to those of a classic Ocean questionnaire, like [that of the Cambridge Psychometrics Center](#).)

Our smartphone, Kosinski concluded, is a vast psychological questionnaire that we are constantly filling out, both consciously and unconsciously.

But it was not just about "likes" or even Facebook: Kosinski and his team could now ascribe Big Five values based purely on how many profile pictures a person has on Facebook, or how many contacts they have (a good indicator of extraversion). But we also reveal something about ourselves even when we're not online. For example, the motion sensor on our phone reveals how quickly we move and how far we travel (this correlates with emotional instability). Our smartphone, Kosinski concluded, is a vast psychological questionnaire that we are constantly filling out, both consciously and unconsciously.

Above all, however—and this is key—it also works in reverse: not only can psychological profiles be created from your data, but your data can also be used the other way round to search for specific profiles: all anxious fathers, all angry introverts, for example—or maybe even all undecided Democrats? Essentially, what Kosinski had invented was sort of a people search engine. He started to recognize the potential—but also the inherent danger—of his work.

To him, the internet had always seemed like a gift from heaven. What he really wanted was to give something back, to share. Data can be copied, so why shouldn't everyone benefit from it? It was the spirit of a whole generation, the beginning of a new era that transcended the limitations of the physical world. But what would happen, wondered Kosinski, if someone abused his people search engine to manipulate people? He began to add warnings to most of his scientific work. His approach, he warned, "could pose a threat to an individual's well-being, freedom, or even life." But no one seemed to grasp what he meant.

Around this time, in early 2014, Kosinski was approached by a young assistant professor in the psychology department called Aleksandr Kogan. He said he was inquiring on behalf of a company that was interested in Kosinski's method, and wanted to access the MyPersonality database. Kogan wasn't at liberty to reveal for what purpose; he was bound to secrecy.

At first, Kosinski and his team considered this offer, as it would mean a great deal of money for the institute, but then he hesitated. Finally, Kosinski remembers, Kogan revealed the name of the company: SCL, or Strategic Communication Laboratories. Kosinski Googled the company: "[We are] the premier election management agency," says the company's website. SCL provides marketing based on psychological modeling. One of its core focuses: Influencing elections. Influencing elections? Perturbed, Kosinski clicked through the pages. What kind of company was this? And what were these people planning?

What Kosinski did not know at the time: SCL is the parent of a group of companies. Who exactly owns SCL and its diverse branches is unclear, thanks to a convoluted corporate structure, the type seen in the UK Companies House, the Panama Papers, and the Delaware company registry. Some of the SCL offshoots have been involved in elections from [Ukraine to Nigeria](#), [helped](#) the Nepalese monarch [against the rebels](#), whereas others have developed methods to influence [Eastern European](#) and [Afghan citizens for NATO](#). And, in 2013, SCL spun off a new company to participate in US elections: Cambridge Analytica.

Kosinski knew nothing about all this, but he had a bad feeling. "The whole thing started to stink," he recalls. On further investigation, he discovered that Aleksandr Kogan had secretly registered a company doing business with SCL. According to a December 2015 report [in The Guardian](#) and to internal company documents given to *Das Magazin*, it emerges that SCL learned about Kosinski's method from Kogan.

Kosinski came to suspect that Kogan's company might have reproduced the Facebook "Likes"-based Big Five measurement tool in order to sell it to this election-influencing firm. He immediately broke off contact with Kogan and informed the director of the institute, sparking a complicated conflict within the university. The institute was worried about its reputation. Aleksandr Kogan then moved to Singapore, married, and [changed his name to Dr. Spectre](#). Michal Kosinski finished his PhD, got a job offer from Stanford and moved to the US.

Mr. Brexit

All was quiet for about a year. Then, in November 2015, the more radical of the two Brexit campaigns, "Leave.EU," supported by Nigel Farage, announced that it had commissioned a Big Data company to support its online campaign: Cambridge Analytica. The company's core strength: innovative political marketing—microtargeting—by measuring people's personality from their digital footprints, based on the OCEAN model.

After the Brexit result, friends and acquaintances wrote to him: Just look at what you've done.

Now Kosinski received emails asking what he had to do with it—the words Cambridge, personality, and analytics immediately made many people think of Kosinski. It was the first time he had heard of the company, which borrowed its name, it said, from its first employees, researchers from the university. Horrified, he looked at the website. Was his methodology being used on a grand scale for political purposes?

After the Brexit result, friends and acquaintances wrote to him: Just look at what you've done. Everywhere he went, Kosinski had to explain that he had nothing to do with this company. (It remains unclear how deeply Cambridge Analytica was involved in the Brexit campaign. Cambridge Analytica would not discuss such questions.)

For a few months, things are relatively quiet. Then, on September 19, 2016, just over a month before the US elections, the guitar riffs of Creedence Clearwater Revival's "Bad Moon Rising" fill the dark-blue hall of New York's Grand Hyatt hotel. The Concordia Summit is a kind of World Economic Forum in miniature. Decision-makers from all over the world have been invited, among them Swiss President Johann Schneider-Ammann. "Please welcome to the stage Alexander Nix, chief executive officer of Cambridge Analytica," a smooth female voice announces. A slim man in a dark suit walks onto the stage. A hush falls. Many in attendance know that this is Trump's new digital strategy man. (A video of the presentation was posted on YouTube.)

A few weeks earlier, Trump had tweeted, somewhat cryptically, "Soon you'll be calling me Mr. Brexit." Political observers had indeed noticed some striking similarities between Trump's agenda and that of the right-wing Brexit movement. But few had noticed the connection with Trump's recent hiring of a marketing company named Cambridge Analytica.

"Pretty much every message that Trump put out was data-driven," says Cambridge Analytica CEO Alexander Nix

Up to this point, Trump's digital campaign had consisted of more or less one person: Brad Parscale, a marketing entrepreneur and failed start-up founder who created a rudimentary website for Trump for \$1,500. The 70-year-old Trump is not digitally savvy—there isn't even a computer on his office desk. Trump doesn't do emails, his personal assistant once revealed. She herself talked him into having a smartphone, from which he now tweets incessantly.

Hillary Clinton, on the other hand, relied heavily on the legacy of the first "social-media president," Barack Obama. She had the address lists of the Democratic Party, worked with cutting-edge big data analysts from BlueLabs and received support from Google and DreamWorks. When it was announced in June 2016 that Trump had hired Cambridge Analytica, the establishment in Washington just turned up their noses. Foreign dudes in tailor-made suits who don't understand the country and its people? Seriously?

"It is my privilege to speak to you today about the power of Big Data and psychographics in the electoral process." The logo of Cambridge Analytica—a brain composed of network nodes, like a map, appears behind Alexander Nix. "Only 18 months ago, Senator Cruz was one of the less popular candidates," explains the blonde man in a cut-glass British accent, which puts Americans on edge the same way that a standard German accent can unsettle Swiss people. "Less than 40 percent of the population had heard of him," another slide says. Cambridge Analytica had become involved in the US election campaign almost two years earlier, initially as a consultant for Republicans Ben Carson and Ted Cruz. Cruz—and later Trump—was funded primarily by the secretive US software billionaire Robert Mercer who, along with his daughter Rebekah, is reported to be the largest investor in Cambridge Analytica.

"So how did he do this?" Up to now, explains Nix, election campaigns have been organized based on demographic concepts. "A really ridiculous idea. The idea that all women should receive the same message because of their gender—or all African Americans because of their race." What Nix meant is that while other campaigners so far have relied on demographics, Cambridge Analytica was using psychometrics.

Though this might be true, Cambridge Analytica's role within Cruz's campaign isn't undisputed. In December 2015 the Cruz team credited their rising success to psychological use of data and analytics. In Advertising Age, a political client said the embedded Cambridge staff was "like an extra wheel," but found their core product, Cambridge's voter data modeling, still "excellent." The campaign would pay the company at least \$5.8 million to help identify voters in the Iowa caucuses, which Cruz won, before dropping out of the race in May.

Nix clicks to the next slide: five different faces, each face corresponding to a personality profile. It is the Big Five or OCEAN Model. "At Cambridge," he said, "we were able to form a model to predict the personality of every single adult in the United States of America." The hall is captivated. According to Nix, the success of Cambridge Analytica's marketing is based on a combination of three elements: behavioral science using the OCEAN Model, Big Data analysis, and ad targeting. Ad targeting is personalized advertising, aligned as accurately as possible to the personality of an individual consumer.

Nix candidly explains how his company does this. First, Cambridge Analytica buys personal data from a range of different sources, like land registries, automotive data, shopping data, bonus cards, club memberships, what magazines you read, what churches you attend. Nix displays the logos of globally active data brokers like Acxiom and Experian—in the US, almost all personal data is for sale. For example, if you want to know where Jewish women live, you can simply buy this information, phone numbers included. Now Cambridge Analytica aggregates this data with the electoral rolls of the Republican party and online data and calculates a Big Five personality profile. Digital footprints suddenly become real people with fears, needs, interests, and residential addresses.

The methodology looks quite similar to the one that Michal Kosinski once developed. Cambridge Analytica also uses, Nix told us, "surveys on social media" and Facebook data. And the company does exactly what Kosinski warned of: "We have profiled the personality of every adult in the United States of America—220 million people," Nix boasts.

He opens the screenshot. "This is a data dashboard that we prepared for the Cruz campaign." A digital control center appears. On the left are diagrams; on the right, a map of Iowa, where Cruz won a surprisingly large number of votes in the primary. And on the map, there are hundreds of thousands of small red and blue dots. Nix narrows down the criteria: "Republicans"—the blue dots disappear; "not yet convinced"—more dots disappear; "male", and so on. Finally, only one name remains, including age, address, interests, personality and political inclination. How does Cambridge Analytica now target this person with an appropriate political message?

Nix shows how psychographically categorized voters can be differently addressed, based on the example of gun rights, the 2nd Amendment: "For a highly neurotic and conscientious audience the threat of a burglary—and the insurance policy of a gun." An image on the left shows the hand of an intruder smashing a window. The right side shows a man and a child standing in a field at sunset, both holding guns, clearly shooting ducks: "Conversely, for a closed and agreeable audience. People who care about tradition, and habits, and family."

How to keep Clinton voters away from the ballot box

Trump's striking inconsistencies, his much-criticized fickleness, and the resulting array of contradictory messages, suddenly turned out to be his great asset: a different message for every voter. The notion that Trump acted like a perfectly opportunistic algorithm following audience reactions is something the mathematician Cathy O'Neil observed in August 2016.

These "dark posts"—sponsored Facebook posts that can only be seen by users with specific profiles—included videos aimed at African-Americans in which Hillary Clinton refers to black men as predators, for example.

"Pretty much every message that Trump put out was data-driven," Alexander Nix remembers. On the day of the third presidential debate between Trump and Clinton, Trump's team tested 175,000 different ad variations for his arguments, in order to find the right versions above all via Facebook. The messages differed for the most part only in microscopic details, in order to target the recipients in the optimal psychological way: different headings, colors, captions, with a photo or video. This fine-tuning reaches all the way down to the smallest groups, Nix explained in an interview with us. "We can address villages or apartment blocks in a targeted way. Even individuals."

In the Miami district of Little Haiti, for instance, Trump's campaign provided inhabitants with news about the failure of the Clinton Foundation following the earthquake in Haiti, in order to keep them from voting for Hillary Clinton. This was one of the goals: to keep potential Clinton voters (which include wavering left-wingers, African-Americans, and young women) away from the ballot box, to "suppress" their vote, as one senior campaign official told Bloomberg in the weeks before the election. These "dark posts"—sponsored news-feed-style ads in Facebook timelines that can only be seen by users with specific profiles—included videos aimed at African-Americans in which Hillary Clinton refers to black men as predators, for example.

Nix finishes his lecture at the Concordia Summit by stating that traditional blanket advertising is dead. "My children will certainly never, ever understand this concept of mass communication." And before leaving the stage, he announced that since Cruz had left the race, the company was helping one of the remaining presidential candidates.

Just how precisely the American population was being targeted by Trump's digital troops at that moment was not visible, because they attacked less on mainstream TV and more with personalized messages on social media or digital TV. And while the Clinton team thought it was in the lead, based on demographic projections, Bloomberg journalist Sasha Issenberg was surprised to note on a visit to San Antonio—where Trump's digital campaign was based—that a "second headquarters" was being created. The embedded Cambridge Analytica team, apparently only a dozen people, received \$100,000 from Trump in July, \$250,000 in August, and \$5 million in September. According to Nix, the company earned over \$15 million overall. (The company is incorporated in the US, where laws regarding the release of personal data are more lax than in European Union countries. Whereas European privacy laws require a person to "opt in" to a release of data, those in the US permit data to be released unless a user "opts out.")

The measures were radical: From July 2016, Trump's canvassers were provided with an app with which they could identify the political views and personality types of the inhabitants of a house. It was the same app provider used by Brexit campaigners. Trump's people only rang at the doors of houses that the app rated as receptive to his messages. The canvassers came prepared with guidelines for conversations tailored to the personality type of the resident. In turn, the canvassers fed the reactions into the app, and the new data flowed back to the dashboards of the Trump campaign.

Again, this is nothing new. The Democrats did similar things, but there is no evidence that they relied on psychometric profiling. Cambridge Analytica, however, divided the US population into 32 personality types, and focused on just 17 states. And just as Kosinski had established that men who like MAC cosmetics are slightly more likely to be gay, the company discovered that a preference for cars made in the US was a great indication of a potential Trump voter. Among other things, these findings now showed Trump which messages worked best and where. The decision to focus on Michigan and Wisconsin in the final weeks of the campaign was made on the basis of data analysis. The candidate became the instrument for implementing a big data model.

What's Next?

But to what extent did psychometric methods influence the outcome of the election? When asked, Cambridge Analytica was unwilling to provide any proof of the effectiveness of its campaign. And it is quite possible that the question is impossible to answer.

And yet there are clues: There is the fact of the surprising rise of Ted Cruz during the primaries. Also there was an increased number of voters in rural areas. There was the decline in the number of African-American early votes. The fact that Trump spent so little money may also be explained by the effectiveness of personality-based advertising. As does the fact that he invested far more in digital than TV campaigning compared to Hillary Clinton. Facebook proved to be the ultimate weapon and the best election campaigner, as Nix explained, and as comments by several core Trump campaigners demonstrate.

Many voices have claimed that the statisticians lost the election because their predictions were so off the mark. But what if statisticians in fact helped win the election—but only those who were using the new method? It is an irony of history that Trump, who often grumbled about scientific research, used a highly scientific approach in his campaign.

Another big winner is Cambridge Analytica. Its board member Steve Bannon, former executive chair of the right-wing online newspaper Breitbart News, has been appointed as Donald Trump's senior counselor and chief strategist. Whilst Cambridge Analytica is not willing to comment on alleged ongoing talks with UK Prime Minister Theresa May, Alexander Nix claims that he is building up his client base worldwide, and that he has received inquiries from Switzerland, Germany, and Australia. His company is currently touring European conferences showcasing their success in the United States. This year three core countries of the EU are facing elections with resurgent populist parties: France, Holland and Germany. The electoral successes come at an opportune time, as the company is readying for a push into commercial advertising.

Kosinski has observed all of this from his office at Stanford. Following the US election, the university is in turmoil. Kosinski is responding to developments with the sharpest weapon available to a researcher: a scientific analysis. Together with his research colleague Sandra Matz, he has conducted a series of tests, which will soon be published. The initial results are alarming: The study shows the effectiveness of personality targeting by showing that marketers can attract up to 63 percent more clicks and up to 1,400 more conversions in real-life advertising campaigns on Facebook when matching products and marketing messages to consumers' personality characteristics. They further demonstrate the scalability of personality targeting by showing that the majority of Facebook Pages promoting products or brands are affected by personality and that large numbers of consumers can be accurately targeted based on a single Facebook Page.

In a statement after the German publication of this article, a Cambridge Analytica spokesperson said, "Cambridge Analytica does not use data from Facebook. It has had no dealings with Dr. Michal Kosinski. It does not subcontract research. It does not use the same methodology. Psychographics was hardly used at all. Cambridge Analytica did not engage in efforts to discourage any Americans from casting their vote in the presidential election. Its efforts were solely directed towards increasing the number of voters in the election."

The world has been turned upside down. Great Britain is leaving the EU, Donald Trump is president of the United States of America. And in Stanford, Kosinski, who wanted to warn against the danger of using psychological targeting in a political setting, is once again receiving accusatory emails. "No," says Kosinski, quietly and shaking his head. "This is not my fault. I did not build the bomb. I only showed that it exists."

Additional research for this report was provided by Paul-Olivier Dehaye.