Everybody lies: how Google search reveals our darkest secrets

What can we learn about ourselves from the things we ask online? US data scientist Seth Stephens-Davidowitz analysed anonymous Google search results, uncovering disturbing truths about our desires, beliefs and prejudices

'Digital truth serum, on average, will show us that the world is worse than we have thought' Seth Stephens-Davidowitz photographed in New York. Photograph: Christopher Lane for the Observer

Seth Stephens-Davidowitz

Sun 9 Jul 2017 04.30 EDT

Everybody lies. People lie about how many drinks they had on the way home. They lie about how often they go to the gym, how much those new shoes cost, whether they read that book. They call in sick when they're not. They say they'll be in touch when they won't. They say it's not about you when it is. They say they love you when they don't. They say they're happy while in the dumps. They say they like women when they really like men. People lie to friends. They lie to bosses.

They lie to kids. They lie to parents. They lie to doctors. They lie to husbands. They lie to wives. They lie to themselves. And they damn sure lie to surveys. Here's my brief survey for you:

Have you ever cheated in an exam?

Have you ever fantasised about killing someone?

Were you tempted to lie?

Many people underreport embarrassing behaviours and thoughts on surveys. They want to look good, even though most surveys are anonymous. This is called social desirability bias. An important paper in 1950 provided powerful evidence of how surveys can fall victim to such bias. Researchers collected data, from official sources, on the residents of Denver: what percentage of them voted, gave to charity, and owned a library card. They then surveyed the residents to see if the percentages would match. The results were, at the time, shocking. What the residents reported to the surveys was very different from the data the researchers had gathered. Even though nobody gave their names, people, in large numbers, exaggerated their voter registration status, voting behaviour, and charitable giving.

> This is one way that researchers collected anonymous data prior to the Internet.

Has anything changed in 65 years? In the age of the internet, not owning a library card is no longer embarrassing. But, while what's embarrassing or desirable may have changed, people's tendency to deceive pollsters remains strong. A recent survey asked University of Maryland graduates various questions about their college experience. The answers were compared with official records. People consistently gave wrong information, in ways that made them look good. Fewer than 2% reported that they graduated with lower than a 2.5 GPA (grade point average). In reality, about 11% did. And 44% said they had donated to the university in the past year. In reality, about 28% did.

Then there's that odd habit we sometimes have of lying to ourselves. Lying to oneself may explain why so many people say they are above average. How big is this problem? More than 40% of one company's engineers said they are in the top 5%. More than 90% of college professors say they do above-average work. One-quarter of high school seniors think they are in the top 1% in their ability to get along with other people. If you are deluding yourself, you can't be honest in a survey.

The more impersonal the conditions, the more honest people will be. For eliciting truthful answers, internet surveys are better than phone surveys, which are better than in-person surveys. People will admit more if they are alone than if others are in the room with them. However, on sensitive topics, every survey method will elicit substantial misreporting. People have no incentive to tell surveys the truth.

How, therefore, can we learn what our fellow humans are really thinking and doing? Big data. Certain online sources get people to admit things they would not admit anywhere else. They serve as a digital truth serum. Think of Google searches. Remember the conditions that make people more honest. Online? Check. Alone? Check. No person administering a survey? Check.

The power in Google data is that people tell the giant search engine things they might not tell anyone else. Google was invented so that people could learn about the world, not so researchers could learn about people, but it turns out the trails we leave as we seek knowledge on the internet are tremendously revealing.

I have spent the past four years analysing anonymous Google data. The revelations have kept coming. Mental illness, human sexuality, abortion, religion, health. Not exactly small topics, and this dataset, which didn't exist a couple of decades ago, offered surprising new perspectives on all of them. I am now convinced that Google searches are the most important dataset ever collected on the human psyche.

Can We Handle the Truth?

I can't pretend there isn't a darkness in some of this data. It has revealed the continued existence of millions of closeted gay men; widespread animus against African Americans; and an outbreak of violent Islamophobic rage that only got worse when the president appealed for tolerance. Not exactly cheery stuff. If people consistently tell us what they

think we want to hear, we will generally be told things that are more comforting than the truth. Digital truth serum, on average, will show us that the world is worse than we have thought.

But there are at least three ways this knowledge can improve our lives. First, there can be comfort in knowing you are not alone in your insecurities and embarrassing behaviour. Google searches can help show you are not alone. When you were young, a teacher may have told you that if you have a question you should raise your hand and ask it, because if you're confused, others are too. If you were anything like me, you ignored your teacher and sat there silently, afraid to open your mouth. Your questions were too dumb, you thought; everyone else's were more profound. The anonymous, aggregate Google data can tell us once and for all how right our teachers were. Plenty of basic, sub-profound questions lurk in other minds, too.

The second benefit of digital truth serum is that it alerts us to people who are suffering. The Human Rights Campaign has asked me to work with them in helping educate men in certain states about the possibility of coming out of the closet. They are looking to use the anonymous and aggregate Google search data to help them decide where best to target their resources.

The final – and, I think, most powerful – value in this data is its ability to lead us from problems to solutions. With more understanding, we might find ways to reduce the world's supply of nasty attitudes. Let's return to Obama's speech about Islamophobia. Recall that every time he argued that people should respect Muslims more, the people he was trying to reach became more enraged. Google searches, however, reveal that there was one line that did trigger the type of response Obama might have wanted. He said: "Muslim Americans are our friends and our neighbours, our co-workers, our sports heroes and, yes, they are our men and women in uniform, who are willing to die in defence of our country."

After this line, for the first time in more than a year, the top Googled noun after "Muslim" was not "terrorists", "extremists", or "refugees". It was "athletes", followed by "soldiers"." And, in fact, "athletes" kept the top spot for a full day afterwards. When we lecture angry people, the search data implies that their fury can grow. But subtly provoking people's curiosity, giving new information, and offering new images of the group that is stoking their rage may turn their thoughts in different, more positive directions.

Two months after that speech, Obama gave another televised speech on Islamophobia, this time at a mosque. Perhaps someone in the president's office had read Soltas's and my *Times* column, which discussed what had worked and what hadn't, for the content of this speech was noticeably different.

Obama spent little time insisting on the value of tolerance. Instead, he focused overwhelmingly on provoking people's curiosity and changing their perceptions of Muslim Americans. Many of the slaves from Africa were Muslim, Obama told us; Thomas Jefferson and John Adams had their own copies of the Koran; a Muslim American designed skyscrapers in Chicago. Obama again spoke of Muslim athletes and armed service members, but also talked of Muslim police officers and firefighters, teachers and doctors.

And my analysis of the Google searches suggests this speech was more successful than the previous one. Many of the hateful, rageful searches against Muslims dropped in the hours afterwards.

There are other potential ways to use search data to learn what causes, or reduces, hate. For example, we might look at how racist searches change after a black quarterback is drafted in a city, or how sexist searches change after a woman is elected to office. Learning of our subconscious prejudices can also be useful. We might all make an extra effort to delight in little girls' minds and show less concern with their appearance. Google search data and other wellsprings of truth on the internet give us an unprecedented look into the darkest corners of the human psyche. This is at times, I admit, difficult to face. But it can also be empowering. We can use the data to fight the darkness. Collecting rich data on the world's problems is the first step toward fixing them.

Extracted from: Everybody Lies: What the Internet Can Tell Us About Who We Really Are by Seth Stephens-Davidowitz, published by Bloomsbury, £20. To order for £17 go to bookshop.theguardian.com or call 0330 333 6846 Free UK p&p over £10, online orders only. Phone orders min p&p of £1.99.. Seth Stephens-Davidowitz will be speaking in London at the Royal Society of Arts on Tuesday and at Second Home on Wednesday

Q&A with Seth Stephens-Davidowitz

'The degree to which people are self-absorbed is pretty shocking': Seth Stephens-Davidowitz. Photograph: Christopher Lane for the Observer

What's your background?
I'd describe myself as a data scientist, but my PhD is in economics. When I was doing my PhD, in 2012, I found this tool called Google Trends that tells you what people are searching, and where, and I became obsessed with it. I know that when people first see Google data, they say "Oh this is weird, this isn't perfect data", but I knew that perfect data didn't exist. The traditional data sets left a lot to be desired.

What would your search records reveal about you?
They could definitely tell I'm a hypochondriac because I'm waking up in the middle of the night doing Google searches about my health. There are definitely things about me that you could figure out. When making claims about a topic, it's better to do it on aggregate, but I think you can figure out a lot, if not everything, about an individual by what they're

searching on Google.

You worked at Google?
For about a year and a half. I was on the economics team and also the quantitative marketing team. Some was analysis of advertising, which I got bored of, which is one of the reasons I stopped working there.

Did working there give you an understanding that helped this book?
Yeah, I think it did. All this data I'm talking about is public. But from meeting the people who know more about this data than anyone in the world, I'm much more confident that it means what I think it means.

What's next?
I want to keep on exploring this, whether in academia, journalism or more books. It's such an exciting area: what people are really like, how the world really works. I may just research sex for the next few months. One thing I've learned from this book, people are more interested in sex than I thought they were.

Interview by Killian Fox