

Rationally Speaking #191: Seth Stephens-Davidowitz on “What the internet can tell us about human nature”

Julia: Welcome to Rationally Speaking, the podcast we explore the borderlands between reason and nonsense. I'm your host, Julia Galef, and I'm here with today's guest, Seth Stephens-Davidowitz.

Seth is trained as an economist, he did his PhD in economics at Harvard. He worked for a while as a data scientist at Google, is a contributing op-ed writer for the New York Times, and very recently published a book called *Everybody Lies: Big Data, New Data, and What The Internet Can Tell Us About Who We Really Are*. Seth, welcome to Rationally Speaking.

Seth: Thanks so much for having me, Julia.

Julia: One thing that really attracted me to your book, Seth, is: first of all, I just love clever experimental design and kind of clever ways of tricking the world into giving up information to us, and your book falls squarely into that category. I have this whole folder of examples of clever experiments and clever studies, on my computer, that I happily added your work to.

More particularly, I ... a big update that has been happening for me in the last few years is that we just can't trust people to honestly and accurately report how they're feeling, and what they believe, and why they do the things they do. If we want accurate answers to those questions, we kind of have to get clever and infer those answers from original sources of data -- like for example, people's Google searches, which is where your research comes in.

My first question for you is why did you first start looking at sources of data like people's Google searches? What was interesting about that kind of data?

Seth: I started when I was in my PhD program. I was kind of little lost and burnt out, I didn't really have a dissertation topic. Then I found Google searches and I just became kind of obsessed with it, because I suspected that people would tell Google things that they might not tell anyone else. You could kind of see what people really thought about various issues, more accurately than asking them.

I just became obsessed, and I started doing this research on racism. That was the first thing I was studying. I was just shocked at how different the results were when you looked at Google searches compared to surveys, and that Google searches seemed to be, in my opinion, more accurate.

So that started me down this whole path which I've been following for five years or so.

Julia: Could you say a little bit more about why self-reports -- people's answers to survey questions about themselves -- why can't we trust those?

Seth: So people lie to themselves a lot. They forget something they did, or ... you know, they might be searching for racist jokes and doing a lot of bad things to African-Americans, but they don't like to think of themselves as racist. So that's one problem.

Then the second problem is that people just have been shown to shade the truth, shade their answers in the direction of things that make them look good, for whatever reason. Nobody knows exactly why. Maybe it's just because it's a habit, people kind of lie consistently in everyday life.

Julia: So it's not-

Seth: You're kind of always trying to make yourself look better, and that habit maybe carries over to a survey.

Julia: Even though they're anonymous on the survey, right? I mean ... I assume most of these surveys are not collecting the person's name and address and everything like that.

Seth: Yeah, they're anonymous, but still sometimes it feels a little weird to people. So even if it's anonymous, they still lie.

I think one big difference in surveys and Google is the survey can never give you an incentive to tell the truth. It might give you an incentive to lie, but there's no incentive to tell the truth. So people will just shade their answers in the direction of what will make them look good. But with Google, you have an incentive to get the information you need.

Julia: Right, right. Good point.

Seth: If you're gay, and you live in a place where it's hard to be gay, you don't have an incentive to tell a survey, "I'm gay." But you do have an incentive to search for gay porn. That would be kind of a classic example.

If you're deciding whether to vote in an election, you don't have an incentive to tell whether or not you're going to vote to surveys, but you do have an incentive to search Google for voting information or voting locations or polling places if you're actually going to vote.

Julia: Right. You mentioned racism as being one of the first topics that you started investigating, and also I can imagine it's a topic where looking at people's Google searches would be especially useful, or add an especial amount of value, relative to the standard social science methodology of asking people about their beliefs. Because racism is this kind of socially charged or socially disapproved of attitude.

Seth: At least it was.

Julia: Yeah... I know, that's why I hesitated.

Seth: Yeah.

Julia: We're recording this episode -- to our listeners, we're recording this episode the week of the Nazi march or rally in Charlottesville, so as I was going back over Seth's book, all the stuff about revealing America's latent or hidden racism was especially salient to me.

What jumped out at you as surprising from looking at the Google search data, compared to common wisdom, or to what other social science research had shown? Actually, sorry ... before you answer that, I suppose I should ask you, where did you get the Google search data? I just can't go online and look at Google search data, can I?

Seth: Yeah you can. There's a tool called Google Trends, where you can type in any search term or any category of searches, and you can see where they're searched most frequently and when they're searched most frequently.

Julia: Great. And it's just sort of a representative, complete ... the statistics are across the whole country, and they're not weighted by where you live or anything like that?

Seth: Yeah, so it's the percent of total Google searches [out of the total].

Julia: Great.

Seth: Some things are a little weird, they make it kind of hard to understand sometimes, and they have a really high privacy threshold... which I figured out how to get around for a while, but now they blocked it. It's kind of a long story but ...

Julia: Because of you? It's the Stephens-Davidowitz Rule at Google now?

Seth: I think, yeah, actually I'm pretty sure it's that.

I worked at Google for a while so... all my data is public, but at Google you can find better ways to find the ideas and then confirm them publicly. But in general, you can learn a lot from Google Trends.

Julia: Okay, so back to my question about what surprised you about racism in America, from this data.

Seth: Okay, well a lot of things that surprised me don't surprise me anymore, but surprised me when I was doing the research.

Julia: You mean given current events?

Seth: Yeah. Wouldn't surprise me now, but when I was doing the research, and people said we live in a post-racial society-

Julia: Right. When was this, what year?

Seth: What?

Julia: What year did you start doing this research?

Seth: I was doing it I think in 2011. Yeah, that was when I started. Obama was elected and everyone was saying we've moved beyond a lot of the really nasty racism in our country's history.

When I started the research, I was just shocked by how frequently Americans were searching for the N word ... not like the "N word," the actual N word.

Julia: The N word but not in quotes.

Seth: Yeah.

Julia: Or not ... yeah.

Seth: I thought it would be ... when I first saw it and how frequent it is, like the time period I was looking at, it was searched as frequently as "migraine" and "Economist" and "Lakers" and "Daily Show." So, not a fringe search.

Julia: Yeah.

Seth: I first thought, oh, rap lyrics. Like, that's what's going on. But the rap lyrics are the version that ends in A, not R. It was basically jokes mocking and humiliating African-Americans, was the big theme of it.

The other thing that was surprising was the location of the searches. Like I would have thought that racism would be predominantly concentrated in the deep South, in Mississippi, in Louisiana, in Alabama, if you think of our country's history. And definitely these were among the places with the highest search volumes, but also among the highest were upstate New York and western Pennsylvania, eastern Ohio, industrial Michigan. The real divide was not so much North versus South, but East versus West.

Julia: Interesting.

Seth: Then it predicted this clear behavior you see back in the day that Barack Obama, compared to other Democratic candidates, did far worse in places that made a lot of racist searches. Like, it was a really, really strong predictor of where Obama underperformed in the 2008 and 2012 elections.

Julia: Interesting. Have you looked at that same correlation with how Trump performed relative to Republican candidates in the past? Did that correlate with the searches?

Seth: Nate Cohn, he's a data journalist at the New York Times, he asked for my data. And he had all this data on Republican primary support, and he found that basically racist searches was the highest predictor he could find of Trump support in the Republican primary. Nate Silver found the same thing.

It's a little harder to compare to other Republicans in the general election because it's compared to the previous election where Obama was black, so it's ... there's a lot going on there, but I think it's pretty clear that racism drove a lot of his support in the Republican primary.

Julia: How confident can we be that that isn't just a result of areas that have declining industries, like in the Rust Belt, people are disillusioned with the current economic regime -- and those areas also happen to be racist, but that's not why they're supporting Trump or dislike Democrats?

Seth: Yeah, so it's just that, basically, Nate Cohn and Nate Silver start controlling for all these other variables. And it was still the racism that was predicting Trump support.

Julia: So even after controlling for things like average income or unemployment-

Seth: Yeah, after controlling for demographics or economics or exposure to trade or anything else, it's ... a big predictor is racism.

Julia: Right. How confident do you think we can be that we're interpreting people's Google searches correctly?

You briefly mentioned considering other explanations for why people might have been searching for the N word, like maybe it's rap lyrics. And you were able to pretty confidently rule that out because of different spellings. But there are a bunch of really interesting findings that you report in your book, and I kept trying to ask myself, could there be other explanations for this, for what people were searching for, other than the obvious straightforward one?

Just to give an example -- I don't think this is in your book, but for the sake of illustration -- if someone searches for the phrase "symptoms of depression," it's kind of unclear whether that's them revealing to Google that they think they might be depressed, and that might give us a window into actual rates of depression, beyond the rate of people who actually seek out help for depression. Or, a different story we could tell is that people have friends who are depressed and want to ... they're worried about their friends and wondering if they should try to help their friends.

How often do you think it comes up that we could be jumping to a conclusion about why people are searching for those terms?

Seth: I think you never know why a particular individual makes a search, but in aggregate, it tells you a lot about patterns. I definitely made a lot of racist searches when I was writing my book.

Julia: Right, but you're probably not typical!

Seth: Yeah, I don't consider myself ... I like to think I'm not racist. I think it'd be different if you got the racist search data and it came back and it's like Cambridge, Massachusetts and Princeton, New Jersey, and like those are the top places, you know, like "Wait, is that just professors doing research or something?"

But when it comes back and it's West Virginia and Louisiana, Pennsylvania and Michigan, I think it's a little more reliable. I think ... when we do kind of get ground truth, you see over and over again that the Google search data correlates with real world outcomes. If you look at the parts of the country that searched for God the most, it's almost perfectly correlated with religious belief, it's the Bible belt and other areas with high rates of religious faith.

I think one of the things that's interesting about that is it doesn't mean that everybody who makes a search with God believes in God. Like, you could search God, you know, proof that God does not exist. Actually I think one of the top searches with God is the "God of War" video game, and that would have to count in the data.

But that's like 5% or 6% of searches, and then it's kind of swamped by all the other reasons you search for God, which, you know, you're looking for God quotes or Church of God or whatever. So that's why it correlates so strongly, I think.

With the health ones, again, when we actually have ground truth, when we have CDC data, Google searches correlate very strongly with the actual health conditions. Also if your friend has depression and they live near you --

Julia: Oh, then maybe that's actually telling us about [about hidden rates of depression].

Seth: Yeah.

Julia: I guess it would still undermine efforts to find correlations. Like, people who search for fashion magazines and then search for symptoms of depression, or something, we couldn't assume that --

Seth: Yes, at the individual level, that would be problematic. But at the community level, that would still frequently work.

Julia: Right. To make sure I understand, the two main defenses you're giving of being able to confidently interpret Google search terms are: First, that when

we are able to check the prevalence of search terms against some objective measure, it tends to show that the search represents indeed what we thought it represented. And two, the patterns that we find for these search terms match what we already expected.

Seth: Oh well, then it wouldn't be interesting, if you totally already expect-

Julia: Well, that's kind of what I was going to say.

Seth: Yeah, but it's not that it's expected, it's just there's not like any other explanation that's... like, if you saw health conditions and they always correlate with hospitals or something. Maybe all the people who are searching health conditions are doctors or something, that would change how you think of it. Or if you saw the racist searches were all in college towns, you could say oh, it's professors. But that doesn't happen. So that kind of gives you some confidence. And with the racism thing, the fact that's what they correlate with where Obama did worse, is I think another proof that it means something.

Julia: What about ... One example you actually did talk about in the book is comparing the rates of people searching for the phrase "I regret having children," versus "I regret not having children." And if I'm remembering correctly, it's much more common for people to search for "I regret having children" than "I regret not having children."

Seth: Yeah, even controlling for how many people have children.

Julia: Oh, interesting. I actually hadn't read or noticed that, so that's actually good to know.

But again -- I don't know how common we should expect this to be, but -- I don't have children, but I'm pretty sure I've searched for phrases like "I regret having children" just because I was curious. I wanted to look at information, examples of other people who had made the choice about whether or not to have children, and were talking about whether they regretted it.

Seth: Yeah.

Julia: I suppose you could say, "Well, I wouldn't expect lot of people to be doing that," but I don't know. It's not obvious to me that you shouldn't expect that.

Seth: I thought for that one, I just thought it interesting the questions people don't ask. People don't ask, "Will I regret having children." Which may be the way, if you were trying to decide whether to have children, you may phrase it that way. But they do say afterwards that they regret having children.

That one, I don't really make too much of that, because it's like a kind of extreme statement. Not that many people. But I think it is just kind of

interesting that any idea, that you get a lot of people who tell Google things that they might not tell other people.

Julia: Yeah. It is also interesting, and you point it out in the book, about how many people use Google as kind of a confessional. They're not searching for phrases like "symptoms of depression," they're searching for things like "I am depressed," or "I am depressed today," you know, which ... I agree, it's harder to interpret that as something other than a statement about the person.

Seth: It's kind of weird, I'm not totally sure what to make of it. I think maybe because people tell Google things that they don't tell other people, they're just in the habit of kind of confiding in Google, that they start typing sentences. But it is definitely ... yeah, "I regret having children" is such a weird thing to type into Google.

Julia: Right. Does anyone type in "Dear Google, I regret having children"?

Seth: No, but people do search for "Google," though, which is kind of weird.

Julia: That might just be ... I bet I've done that though. I think that's probably just muscle memory, like I forget ... I'm not paying attention, I've forgotten I'm actually on Google and I'm typing as if I'm typing something in the URL window.

Seth: Yeah.

Julia: I don't know, maybe it's just me.

Seth: Yeah. But do you actually think you would type in, if you were trying to decide what happens, regretting children, you think you'd type in "I regret having children versus," like "Do people regret having children," or "Are children a good decision?"

Julia: Yeah, I don't know-

Seth: A full sentence, you're basically saying ... that starts with "I," you're basically saying that that's you.

Julia: I agree it's the sort of natural straightforward explanation. It definitely feels more common sense. I just ... I keep getting burned with psychology, social psychology where the common sense assumption about what's happening just doesn't get borne out by the data. And so I've been trying to cultivate this extra layer of wariness of assuming that my interpretation must be correct if it sounds right.

Seth: Yeah. It's kind of also, Google search data can be as good as we want it to be ...

Julia: How so?

Seth: Because you could ultimately follow individuals over time. I just had aggregated anonymous data, but you could know people's internet behavior over time, and then you'd know if they actually had children and then you'd be like, oh okay, that's the first time --

Julia: Oh, when you say you could, you mean Google could?

Seth: Just in theory, I mean. That data exist. It's not made available, but ... you know, or depression, you can probably figure out whether someone's actually making a search about themselves, if you had all their information based on their online activity.

But sometimes people search "panic attacks" at like 3 A.M. or something, and I think you're pretty sure at that point that that person's actually having a panic attack, even if not everybody who searches for panic attack is having a panic attack. There definitely can be clues if you kind of dig down deeper into the data, that could tell you with more confidence what's actually causing that search.

Julia: You talked about a kind of validation of the Google search results by comparing them -- aggregate search terms for disease, for example, comparing them against rates of the disease.

Is there any way to validate Google search terms as a metric for other things at the individual level? Like ... I don't know, again, it seems pretty intuitive that people who search for the N word online or for N word jokes are more likely to be actually racist, but is there any way that we could in the future -- I understand your research is pretty new, but -- some way that we could theoretically check to see if people who search for those things are more likely to behave in a racist way? Or more likely to ... I don't know, judge people's resumes differently if they're black versus white, even if it's the same resume, things like that? Does that seem worth doing?

Seth: Not at the individual level yet, although you know, maybe someone will kind of ... you have to, at this point, get people to volunteer to give you their search data and then do some experiments based on that. I guess you could do it.

But they already are comparing the aggregate data on racism to various offline behaviors, the Obama voting pattern was one, and then the Trump support. And then recently people have found that wages, that places with bigger white-black wage gaps have higher Google [search rates], and this seems to survive a lot of controls. It does seem like at least in aggregate, this is predicting something pretty important.

Julia: Yeah. One interesting thing that you can do with this kind of data, that you might have mentioned briefly but we haven't really talked about yet, is the

temporal component. You can look at how searches change in response to certain events or over time in certain areas. One cool example you talk about in the book is the effects of Obama's speech after some terrorist attack. I could explain it, but why don't you talk a little bit about Obama's speech and what you were able to learn from people's searches during and after it?

Seth: Yeah, so this was with Evan Soltas, who I say is a "scholar" at Princeton, but he's actually like a junior. He's like a total prodigy, you're going to be seeing his name in the future. If this is the first time you hear him, he's going to be big in the future.

Julia: Good to know. We'll see searches for Evan Soltas go up.

Seth: Yeah. We're doing this research on Islamophobia when ... it's not really a phobia, it's like rage against Muslims. You kind of see there are a lot of maniacs who make searches like "I hate Muslims," or "Kill Muslims," or ...

Julia: That is such a weird thing to type into Google, just like, "I hate Muslims." What do you expect Google to give you?

Seth: I think it goes into this idea that people make ... when they're emotionally charged, make kind of statements that aren't necessarily, it's not clear what they're hoping to get from Google. Again, these guys aren't really necessarily ... I assume guys -- and gals, I guess -- are not exactly totally sane, so they're kind of just expressing some sort of rage.

Julia: Not sane in that moment? You're not, you know, suggesting that they're mentally ill.

Seth: Yeah, you'll also see it at the highest, like 3 A.M., which kind of also gives you a sense of what these people are ... maybe they're not sleeping. But then they also predict ... even though they are these weird searches, like "What does it even mean?" they actually predict very strongly hate crimes against Muslims. So when these searches are high, there tend to be more hate crimes committed against Muslims.

But then we were analyzing these searches after the San Bernardino attack, and after the San Bernardino attack, there was an explosion of these searches. Like the top search with the word "Muslim" was "Kill Muslims."

Julia: Wow.

Seth: A few days afterwards, Obama gave this speech where he was kind of trying to calm down this mob. I think a lot of people had realized that something had gone out of control in in Americans' attitudes towards Muslims, and he gave this speech that was nationally televised and got a lot of attention.

It was kind of a beautiful speech, kind of classic Obama, got rave reviews from all the serious sources, from the New York Times, the LA Times,

Newsweek, saying how beautiful a speech it was. And he talks about how it's our responsibility as Americans to not give in to this fear, to kill the freedom, to not judge people based on their religion.

Evan and I were actually working on a New York Times column on this topic during that week when Obama gave the speech, and we're like, "Oh, let's see if this helps calm things down." Our idea was that it probably would have. People thought it was a great speech, everyone else seemed to think it was a great speech. We went to the data and we said, "Was there a big drop in these really nasty searches about Muslims during and after Obama's speech?"

And we see that not only did the searches not drop, they didn't stay the same. They definitely shot up and stayed up afterwards. More searches for "Kill Muslims" and "I hate Muslims" and "No Syrian refugees" and "Muslims are evil." So it seemed like, everything Obama was doing was actually backfiring.

But then at the end of the speech, Obama kind of gives this one line that seemed to have a different effect, where he said that, "We have to remember that Muslim Americans are our friends and neighbors. They're our sports heroes. They're the men and women who will die for our country."

Then basically 30 seconds after he says this, you see for the first time the top descriptor of Muslim on Google was not "Muslim terrorist," or "Muslim extremist," or "Muslim refugees." It was "Muslim athletes," and followed by "Muslim soldiers." These kinda kept the top two spots for many days afterwards.

Julia: Interesting.

Seth: What we suggested in our Times piece, Evan and I, was that basically if you want to change peoples minds, kinda calm an angry mob, you don't want to lecture them about things they've been told a thousand times, and tell them what they should do, and what is their responsibility. But maybe provoke curiosity, change how they think about this group that's causing them so much rage.

Then -- we published this in The New York Times, then -- I think it's not crazy when you write a New York Times column to think that people in high places read that. Perhaps in Obama's government.

Because Obama a couple weeks later gave another speech in a Baltimore Mosque, and again it got a lot of attention. Again it was on national TV. But this time he basically stopped with all the lectures, and the sermons.

He doubled down on the curiosity strategy, where he talked about how Muslim Americans built the skyscrapers of Chicago, and how Thomas Jefferson had a copy of the Quran. Then you see after this speech most of these anti-Muslim searches actually dropped.

So it did seem ... like, obviously I'm not gonna take from two speeches that we learned how to end hatred. But I think it is suggestive, and does suggest that we could use some of this data to turn something seemingly as insane as how to calm an angry mob into a real science.

Julia: That is really interesting. Is there any way to tell from the data that we have whether ... I mean, there's two stories that you could tell. One is that people who hadn't previously been angry at Muslims were made angry by Obama's speech -- the first part of it, anyway, of his first speech. So they started searching for, "kill Muslims."

Or, you could tell the story that the same people who were angry at Muslims are angry at them still, so Obama's speech didn't help. He's reminded them that they dislike Muslims, and now they're doing more searches. So is there any way to tell whether it's more searches from the same people, like the same IP addresses, or searches across a broader range of people?

Seth: In theory, yes, because that data exists. But not with the data that is publicly available now.

Julia: Got it.

Seth: You could look at areas, which we haven't done. You could look at areas, you know, since there is significant variation in how frequently these searches are made in different areas. You could say, Is it areas that didn't previously have these attitudes, versus areas that had these attitudes in big numbers? But you couldn't, at least now, do it at the individual level.

I'm kinda hoping that because of the power of this data, like, there will be more support for, while protecting peoples anonymity, giving some of the data at the individual level. Because I do think it is really, really powerful.

Julia: Tell the truth -- you're secretly hoping that some hacker is just going to release a ton of search data that you can then use for research, right?

Seth: I am. And also because it'll be embarrassing for everybody else, but it wouldn't be embarrassing for me. 'Cause I can credibly say every search was for research purposes.

Julia: Oh, that's so true. Oh, man. I should totally write a book about weird sexual preferences and I'd just be covered for life.

Seth: Yeah. Well actually, AOL ... I have used this data a little bit in my research, and I talk at one point in the book about how AOL released their data anonymously, in aggregate, to researchers. It was a huge disaster, which is one of the reasons that-

Julia: Because it could be de-anonymized right?

Seth: Yeah, because they just didn't think twice, like, some mid-level employee just gave it to researchers. Just like, "Yeah, here, this sounds interesting."

Julia: Right.

Seth: Then someone would search their address and their name, and then like "herpes symptoms" or something.

Julia: Right. Oh, God.

Seth: People are figuring all this out. Yeah. So now they're, I think rightfully, very cautious. But I think there are ways to do this, in ways that would protect people's anonymity but still help with the research.

Julia: Maybe Google could hire some really good hackers to try to de-anonymize the data, and only when the hackers failed would they release it to researchers. Oh, but they'd have to kill the hackers afterwards...

Seth: They do that with a lot of their products, that's the way they test a lot of their products. Yeah.

Julia: Oh, nice. What was I gonna say? Oh, just one last thing on people's reactions to Obama's speech thing, and people's reactions to different strategies to get them to be less racist, or less hateful: The thing that motivated my question, about "Is it the same people, or a broader base of people, doing racist searches?" ... was that I have this worry, or intuition -- this hunch... People tend to look at the most extreme members of society and what happens with them to determine what is the right outreach strategy. Or the right rhetorical approach. But it might just be the case that the strategy that works best for most of the population works worst for this most extreme group of people.

So, for example, it might be the case that the majority of the population is shamed into trying to not be racist by speeches like this. Not necessarily because they like Obama and want to live up to his ideals, but because they think that Obama represents society's ideals, and they don't want to be judged a bad person by society.

Then there's the small minority of people who react very strongly against that. They're maybe the people who visit 4chan and who are making the alt-right a thing now. And they're gonna react really negatively against that, and sort of do the opposite of what the prestigious societal leaders tell them to do. And we just can't have it both ways, you know? We have to choose whether we want to try to de-radicalize the alt-right, or to shame the majority of society into not being racist. That kind of thing.

This is just a hunch.

Seth: Yeah, it might be. I mean, I think, for the hate crimes thing, with what was going after the San Bernardino attacks and what Muslim Americans have been

experiencing after terrorist attacks, I think they probably will be more concerned with the extreme fringe members. The people who tend to shoot Muslims, or attack Mosques, or terrify Muslims. But, I think that was more the goal of Obama's speech at that point. But, yeah, it's certainly possible.

I think in general that's right. That is something that probably we don't do enough of in research, which is -- we usually like to test whether something is effective or not, on average. Clearly things that are effective for one group may backfire for another group.

Julia: Right. Do you think that it's good on net for it to become common knowledge that ... just how frequently people search for these things that are generally considered shameful or secret? Like secret prejudices, or secret fears about their body, or things like that?

Seth: I don't know. Initially I thought it was ... So I think the secret body one is definitely -- you know, like secret insecurity -- I think is generally good. Because I think when people learn that they're not alone in their insecurity... You know, I talked a lot about men's bodily insecurity and their focus on man boobs or whatever. Which is kinda, it's actually a serious issue.

Julia: No, right.

Seth: There are a lot of serious issues that I talk about in the book. Where I think making people aware of how common these fears are will make them feel less alone.

But like, you could say that the racism thing is like, "Oh, why should I feel so secretive? Why should so bad in my racism when there are all these other racists out there?"

Julia: Yeah. So that, I was kinda being a little tricky by lumping those two examples together as if they were examples of the same thing. Because I feel exactly the same way. People's insecurities, or weird sexual picadillos or whatever, it seems good for it to be known that -- you feel weird, but you're not actually that weird.

But when it comes to racism I'm really torn. Or other sort of socially damaging attitudes. Because on the one hand-

Seth: Well even the sex-

Julia: Yeah, go on?

Seth: Even the sexual stuff, it's not clear that ... We have this idea that, you know, that like there's something good about a large number of people having it. Like if five percent of men are gay then men should not be embarrassed -- and I don't think they should, to be gay.

But like, if you have a sexual preference that one in a hundred thousand people have, I don't think, necessarily, that means you should be embarrassed of it. That's

definitely something that this data would also reveal. Things that maybe you thought were more common aren't actually as common.

Julia: Oh, I see, it could go the other way. Right.

Seth: Like, yeah, I would say pretty much ... Everyone always assumes that everything I research I'm talking about myself. But everything I discovered was not me at all. All my insecurities, it turns out, are just totally weird.

Julia: That's so funny. Oh, wow.

Well, so about the racism one: The way I've been thinking about it recently, even before reading your book, just looking at Trump and sort of the way that the alt-right has become so much more main stream than it was five years ago... It doesn't seem to me, necessarily, that the country is becoming more racist. But it does seem to me that Trump and the whole discourse around Trump is creating *common knowledge* about the racism of this country.

I've just been really torn about whether that's good or bad, because on the one hand you could tell a story where it's good, and that, well -- at least now our country knows. We know what we're up against, maybe now people will sort of take us seriously when we say that it's important to fight racism, 'cause it's become much more explicitly a problem. Or, like, the problem has become more explicit.

But on the other hand you could tell a story where it's bad. Because, let's say racism is really common and a lot of people, maybe even a majority of people, are actually racist and search for N word jokes online. They're still not necessarily going to feel emboldened to push publicly, as a group, for racist policies, if they don't realize that a lot of people are also racist. Even if each individual person knew that racism was common, they still wouldn't be emboldened, unless they knew that other people also knew that this was a common belief.

It's only when you get to this third stage, of common knowledge, where each individual racist person knows that *everyone else knows* that racism is common. And it's there where you start, that's like the spark that makes those individual people feel like they'll be totally fine and protected if they start pushing these attitudes in public.

That feels like the stage that we're getting to with Trump. It also feels like the kind of thing that -- apologies, but -- research like yours could sort of help create.

Seth: I mean, I'm gonna be honest, I just really interested in things and don't think of the third level, of common knowledge.

Julia: I feel you, man. And I'm not pushing for censorship at all. Like, I-

Seth: I also think it does get hard, once you start having those questions in your mind. Like, it becomes really hard to even do research.

But I do take your ... I think the other one, I think antisemitism is another one where ... it's not even that antisemites don't know that there are other antisemites with them. I've done a lot of research on Stormfront and a lot of the white nationalist movement, and I think a lot of people -- it kind of tends to be younger people. More frequently men, who are just kind of unhappy with their life, and looking for something to latch onto. I think literally just hearing the word Stormfront, or hearing the phrase, "Jews create problems in society," can allow that to fill the vacuum instead of something else.

Julia: Oh, I see. So you're suggesting that the risk is not the thing I was describing, of people who are already racist feeling emboldened to be public about it, or push for racist policies publicly. You're saying people who are sort of at a tipping point, and they could go either way -- if these views are more common or mainstream, or they're more likely to encounter them, that could tip them into being racist when they otherwise wouldn't?

Seth: Yeah, I think with antisemitism that's the fear. 'Cause in the United States it's not, like -- at least two years ago, it just wasn't on people's minds. I don't think many people were hearing these conspiracy theories.

Julia: Yeah.

Seth: Now they are, more. You actually do see them in Google search data, where like, Steve Bannon gets in the news, then you see people in Montana searching for Steve Bannon, then they're searching for white nationalism, then they're searching for Stormfront, then they're searching for "Jews are evil."

Julia: Oh my God.

Seth: So it's like, I think it just wasn't on their mind at all, and then by being told it, it is on their mind. So, yeah, this is all incredibly explosive and dangerous.

Julia: Yeah.

Seth: I do think we can use this data to research and understand better what causes it. But, maybe I should leave my research for academic journals that nobody reads and not The New York Times.

Julia: Just write really boring, really dry articles.

Seth: Yeah.

Julia: Just, the last thing on that point. You're probably more familiar with these studies than I am. But, I know there's a whole body of research in, I think, behavioral economics studying how social proof works as an influence tactic. What they find is that it often backfires.

Like when we try to urge people to stop some behavior, often the way we intuitively try to phrase it is -- well, I'll take a simple example: "God, you know, everyone's been

failing to rinse their dishes off before they put them in the dishwasher. Please stop that, because then the dishwasher can't clean them." You might think that such a statement would cause people to rinse their dishes off before they put them in the dishwasher, but actually what it does is it backfires. Because the part that people focus on is, "Oh everyone's doing this? Everyone's not rinsing their dishes off? Well then, I feel like it's okay for me to not rinse my dishes as well. Because apparently this is a more common thing than I realized, so I don't have to feel so guilty about it like I used to."

Seth: Yeah. It's really interesting. I think we definitely can use this data to understand all these effects more, like social proof and backfiring. But yeah, definitely just as far as making ideas available to the masses, it's obviously complicated what helps and what hurts.

Julia: Yeah. Anyway, I'll leave you alone on this topic. I'm sympathetic, it's a tricky question, and I don't favor censorship, so. I'm not telling you to stop doing your research.

Let's instead talk about political polarization. So this was another really interesting part, for me, 'cause I've read a lot about the filter bubble theory. That the internet and social media are making us all more politically polarized.

Because, well, for one thing there's just so many different sources we could choose to consume on the internet. So because of confirmation bias, and motivated reasoning, we have more opportunity to seek out the sources that agree with us and validate our preexisting beliefs on the internet than we used to. Then secondly, tech companies like Google or Facebook are making the problem worse by customizing our search results, or customizing our Facebook news feed with an eye to results that we're going to like, click on and read. And that's gonna be the stuff that supports what we already believe. So they're sort of making the filter bubble problem even worse.

You talk about some research in your book that contradicts this theory. How does that go?

Seth: Yes, but it's not my research, it's Jessie Shapiro and Matt Gentzkow. They've basically been studying exposure to different political views offline and online. So how frequently do you watch TV, watch news with different viewpoints? How frequently do you have friends with different viewpoints? How frequently do you have work colleagues with different viewpoints? How frequently are you exposed to different viewpoints online?

They say that contrary to this conventional wisdom, you're actually more likely to be exposed to viewpoints online than offline. There are kind of a couple reasons for this. One is that despite this idea that there's kind of this long tail, and there are all these random websites, and you can find what ever information you want... The vast majority of people get most of their news from only a few sources. You know, Yahoo News, I think AOL News still maybe, or a couple other ones.

Julia: Really, people use Yahoo News?

Seth: Yahoo News is one of the big ones. I don't even know.

Julia: It just goes to show I'm in a bubble 'cause I had no idea people use Yahoo News.

Seth: Yeah, all this research on the internet just tells me how good I ... I thought it was gonna make me feel better, but I'm like, weird in all dimensions.

But yeah, so there are like four main sites that everyone goes to. The other reason is that, like, social media, that's kinda considered the biggest cause of this filter bubble. That people just talk to their friends. But even that's not quite as extreme as people think, 'cause on Facebook ... This might be different on Twitter, which I haven't studied yet. But on Facebook you tend to associate with your weak ties. So you're not just friends on Facebook with the people you hang out with offline, but you're also friends with uncles you haven't talked to in five years, and high school acquaintances you haven't talked to in 20 years. These people are much more likely to have different viewpoints than you do.

I think one of the big points to that research is that we experience filter bubbles offline. Like a huge number of the people we come across in our offline lives, share our political views. So our friends share our political views, our family members share our political views, our colleagues share our political views.

Julia: That's been going up over the years hasn't it?

Seth: Yeah it has. Our neighbors share our political views. So, like, even though it's true that people online, we tend to associate more with people online with our political views, than without our political views, it's not more extreme than the phenomenon offline.

They also found, recently, that the biggest increase in political polarization has been amongst elderly people -- who are the least likely to use social media. And the least likely to use the internet frequently. So it doesn't seem to be caused by the internet.

Julia: That is definitely interesting, and it throws a wrench in my common sense model of the world. I almost wonder if we need a two-factor model -- where there's, maybe, use of social media on the one hand, and then consumption of, like, talk radio on the other hand. Maybe the consumption of talk radio explains the polarization of older people.

I don't know, I'm just spitballing here. But it's not obvious to me that—like, the fact that older people use social media less, and are more polarized... assuming that is true, it doesn't definitively refute the theory that social media could be, all things equal, controlling for talk radio, could be making things worse.

Seth: Yeah, there's a lot going on. But I think there are other points in general, like there's huge polarization offline, I think is definitely true.

Julia: Yeah, that is true, and it's getting worse. So it's gotta be part of the solution – sorry, part of the explanation, rather.

Seth: Yeah.

Julia: But, okay. I'm just talking, not about research, but about my own anecdotal observations and impressions. One thing that social media seems to be worsening is that ... yes, in real life I'm going to sometimes -- not super often, but sometimes -- encounter people who have different political opinions from me. But probably our interactions are going to be kind of polite and friendly, because in general, people in real life are polite and friendly to each other.

But online, people's interactions with people with differing opinions are much less likely to be polite and friendly. Because the internet, you know, it emboldens us to not be polite and friendly. We don't really feel like those other people are people.

So it seems quite plausible to me that there's something about the anonymity of the internet -- and those weak ties that we're encountering, you know, saying things that we disagree with -- that could be making polarization worse. Making people feel antagonized by, and antagonistic towards, people with differing views, that they didn't previously feel. Does that contradict the data?

Seth: Yeah, that definitely... Well, Facebook, it wouldn't be anonymous. It would probably be more anonymous if you just randomly come across someone who you're not gonna see again than a weak contact on Facebook.

Julia: So, "anonymous" isn't quite the right word. But still I think, even on Facebook, with people whose names you can see, even people whose identities you know, something about the Internet medium makes people more likely to be rude, I've found.

Seth: That's possible, I don't know. I mean, I remember, I've definitely been in, in person situations that have been pretty rude. I'd be interested to see the data on that. I remember in high school, which was the last time I was with a lot of people with opposing political views, for me it was pretty wild, and rude, and obnoxious, and screaming. So, I don't know.

Julia: Okay. Maybe it's my politeness bubble then.

Seth: I don't know. Yeah, I think, it definitely does seem like the political conversation on Facebook is unusually hostile. But, I don't know.

Julia: Well, one other thing I wanted to make sure I ask you before we have to wrap up, is a problem in general -- with research of any kind really, but especially with big data. It's the problem of data mining: So, you report all of these examples of fascinating correlations. And often there's a very plausible story for that correlation and it's hard to come up with other plausible stories. But still, in the back of my mind as I'm reading these findings, is "How many other correlations did Seth, or the researchers, look for?"

Like, if you find, for example -- this is just a made up example -- that people who search for fashion magazines, or read fashion magazines online, are also especially likely to search for the phrase, "I'm fat." That could be presented as evidence that reading fashion magazines makes people self conscious or insecure. But it could also be the case that there were a lot of other correlations that the researchers searched for, like searching for the phrase, "I'm overweight," or, "I'm ugly," or, "My hair is ugly," or something like that. Maybe there weren't any of those correlations, but you find one for the phrase, "I'm fat," and present that as evidence of this effect.

And that's just always a problem in research. But it just seems especially like there's so many different things you look for in this data, and in big data in general. Is this a problem you're sort of trying to mitigate? Do you like, pre-register your studies in any way, or keep track of the correlations you search for, or anything like that?

Seth: I definitely try to mitigate it. I agree it's a huge problem, so, you know -- the things I present in the book were things that I thought survived. I talk about how the top question women have about their husbands is, you know, if you do the phrase, "Is my husband," the top way to complete that is, "Is my husband gay?"

Julia: Yeah. Wow.

Seth: They could be like, if they do it in other ways. Like, maybe, when they're searching whether he's depressed, they're like, "Signs husband is depressed." That's much more common than, "Signs husband's gay." But if you actually put together all the different variants -- you know, it's not exactly the numbers I put, where it's eight times more likely to be depressed. But maybe if you put them all together it's like seven point eight, six times more likely. So I felt like that one, you know, held up. Or like, I talk about how the top search with "My husband wants," in India is, "My husband wants me to breastfeed him."

Julia: Yeah, that one really threw me for a loop. No offense, India.

Seth: Yeah, which really throws people for a loop and is pretty wild. Especially since nobody talks about, like nobody acknowledges it even after this research. As I found, but that one could be like, is that just one way to phrase it? But, if you do all kinds of different variants of the same thing, like if you do, "Tips on breastfeeding." Like in pretty much all country's in the world, not surprisingly, 99% of searches on "Tips to breastfeed" are tips to breastfeed a child. But in India they're about equally split between tips to breastfeed a child and tips to breastfeed a husband.

Julia: Oh, my God. Wow.

Seth: So I definitely do, I like to think that the things behind my research are thought through. The other thing that Google does is they have categories now. So if you're looking up anxiety, they'll categorize a whole bunch of anxiety searches, and related searches. So "anxiety symptoms," "anxiety health," and "anxiety what do I do." Anxiety and anxious, whatever. They'll put them all together in a basket of anxiety. So I think that helps a lot of people.

Julia: That sounds terrifying. A basket of anxiety.

Seth: Yeah. It's terrifying, but it does help with the cherry-picking, I think. Because they've kinda done all the work behind the scenes to make one big category.

But I think, you know, it's definitely an issue. I think in general you want things that work out of sample, and hold up to other researchers. So, definitely when you're reading something, if it's a first pass and hasn't been reproduced 100 times, I think some skepticism is definitely warranted.

Julia: Hey, Seth, maybe you could make available your search history for the search terms! So that people could see what you searched for, and what different combinations your tried, and so on and so forth.

Seth: Yeah. They'll see what a liar I am, and how much cherry picking I do.

Julia: And you could just say, well, you know, everybody lies. It's right there in the title. You know, I told you up front.

Well, we're just about out of time. So, before we close, Seth, I want to invite you to give the Rationally Speaking pick of the episode. That's a book, or a blog, or a website or something that has influenced your thinking in some way. What's your pick for this episode?

Seth: Well, actually I had a meta point about this pick.

Julia: Oh, sure. As you know, I'm all about meta points.

Seth: An hour and a half before our conversation, when you told me that I need to pick something that influenced my thinking, I started looking through my Kindle, and I realized that all the things on the top that I've been reading recently, and that have hugely influenced me, are all books I'm extremely embarrassed about and would never announce to the world.

Julia: Amazing, I love it.

Seth: That kind of goes to *Everybody Lies*, right? So I had to go through, like, getting through all this cheesy self help things, and-

Julia: So we should really all share our, you know, Kindles and our Netflix watching history and everything. So we wouldn't be embarrassed.

Seth: Right. I think that might be more useful than asking for a pick, which inevitably gets me to find something that maybe people haven't heard of but makes me sound like I know something other people don't, and intellectual, whatever.

Julia: Yes, I should just ask my guest on the air to open up their Kindle and randomly pick a book and tell me what it is.

Seth: Yeah.

Julia: Anyway, sorry. So what is your filtered, you know, respectable-ized pick?

Seth: I'm gonna go with Steven Pinker's "The Better Angels of Our Nature." I know everybody knows this book, but the reason I'm recommending it is because I hadn't read it actually. I just knew about it and I knew the argument.

I didn't really see anybody come back against the argument. So I thought, okay, like, the media exaggerates. You know, the media loves to talk about individual crimes. They make us think that we have this huge crime problem, but really it goes down over time. I just thought that, like, two sentence ... or whatever, that summary was enough and all you need to know.

Julia: Right.

Seth: But recently I read the whole book and I was kinda blown away by how thoughtful it is, and how smart it is. The whole idea of the civilizing process, how everybody's impulses are kind of ... that people have an inherent tendency to just follow their impulses. You have to basically stop people from doing that, in all dimensions. That's kinda how we got violence down, and one of the reasons -- I don't know if I totally believe this, but -- one of the reasons Pinker thinks crime rose in the 1960s was because the hippies kinda stopped this whole idea that you have to not follow your instincts. So, I just found it very smart and very interesting, and worth a 700 page read.

Julia: Nice.

Seth: Even if you know the two sentence summary.

Julia: So you updated about human nature from the book, and civilization, I imagine. Did you also update about your ability to understand books based on their summaries in the media? Or just Pinker, is he an exception?

Seth: Pinker in general, I think. I've kinda re-read all his books recently and I've always had the response that they're a lot better than I had realized. I think at first I either got the summary, or skimmed them while I was reading my self help and then more recently I've read them more seriously, and been very impressed by them. But I don't know if that's a general principle.

Julia: I mean one general principle I've been noticing is how often people will claim, like, "So-and-so wrote in his book..." and then they'll say something offensive. But the actual book makes a much more nuanced point that explicitly disavows the offensive interpretation. But it becomes common knowledge, because very few people actually read the book, so they just repeat what they heard other people say was in the book.

Seth: That definitely, yeah.

Julia: Anyway, excellent. We'll link to "Better Angels of Our Nature" as well as to your own book, "Everybody Lies: Big Data, New Data and What the Internet Can Tell Us About Who We Really Are." Thanks so much for joining us Seth. It's been a pleasure.

Seth: Yeah, thanks so much Julia.

Julia: This concludes another episode of Rationally Speaking. Join us next time for more explorations on the borderlands between reason and nonsense.