Renée DiResta: It started with Kermit memes. A lot of Kermit sipping tea memes, you know, kind of Kermit commenting on Miss Piggy. It's actually kind of raunchy, the post is very irreverent. Then one day there's a post of Homer Simpson and it says something like, I killed Kermit or Kermit's taking a break or Kermit's gone now this is my page.

Tristian Harris: That's Renée DiResta, one of our nation's leading experts on information warfare. In 2017 the Senate Intelligence Committee asked Renée to investigate Russian attempts to manipulate voters through social media by handing her datasets from suspicious social media accounts. She started piecing together what Russian agents posted to these accounts from day one and what's the first thing she sees, images of Kermit the frog and Homer Simpson. She's mystified.

Renée DiResta: Is my dataset broken or my number is pointing to the wrong things? What the hell is going on here?

Aza Raskin: What's going on here is one of the least understood aspects of how disinformation campaigns work. Russia's campaign, for instance, didn't necessarily begin with a masterful manipulation of voter sentiments. Many accounts make no overt reference to politics at all. Instead, they posted content that was eminently likable. They cycle through beloved cultural icons like Kermit the frog, Homer Simpson or you Yosemite Sam, because their goal was deceptively simple, rack up followers.

Renée DiResta: And so you see them doing the hashtag follow back and all... there's 20 hashtags, 30 hashtags per post at the beginning.

Tristian Harris: Like, follow us back.

Renée DiResta: Follow us back basically. So just trying to get new followers.

Tristian Harris: They're trying and failing, but they learn from their mistakes, because the memes keep evolving. They're getting stickier until one day salvation arrives.

Renée DiResta: I want to say maybe even 900 memes before they finally got to what this was, which was the Army of Jesus page. Many people have seen Army of Jesus content because the Senate and others have shared it. It was pictures of Jesus often with a Maga hat on.

Tristian Harris: It's tempting to laugh at these bizarre memes. I mean, Renée has laughed at a few herself, but she argues that these memes are no laughing matter for all their crudeness, they're actually really sophisticated. They represent an evolutionary leap in propaganda, perfectly adapted to our times.

Renée DiResta: One of the things that I talk about a lot, particularly with Russia and when we talk about disinformation campaigns now, I hear a lot like, "Oh, it's just some stupid memes." And it's interesting to me to hear that because I'm like, well, you know, they were running the same messages in the 1960's in the form of long-form articles. And that's because in the 1960's people were reading long-form articles or they were listening to the radio. And so you would hear propaganda on the radio. So the propaganda just evolves to fit the distribution mechanism and the most logical information, kind of the reach of the day. And that's why you see propaganda kind of evolving into mimetic propaganda. So in a way they should be using memes, in fact, that is absolutely where they should be. And it's interesting to hear that spoken of so dismissively.

Aza Raskin: Today on Your Undivided Attention, we're going to take this new form of viral propaganda seriously. We'll ask Renée how bad actors can craft a message that seizes our attention and plays on our fears and grievances.

Tristan Harris: We'll see how they game social medias algorithms to promote their inflammatory ideas, and more importantly, drown out the voices of reason. We'll see how a few users can overwhelm an online forum and create the illusion of consensus, and we'll see how fringe ideas as they enter into mainstream, can lead to world consequences on a scale that we're only beginning to comprehend. Kermit was only the opening shot. I'm Tristan Harris...

Aza Raskin: And I'm Aza Raskin.

Tristian Harris: And this is Your Undivided Attention.

Tristian Harris: Renée, thank you for, for coming on the podcast.

Renée DiResta: Thanks for having me.

Aza Raskin: Renée, I am so excited to talk to you, because I think you are sitting on top of one of the most interesting views of how do our brains work? How do our cognitive biases work, our social biases work, so that not just our behavior is manipulated, but that our identity's manipulated so that we end up becoming useful idiots. I just love to hear your overarching frame of like what's going on.

Renée DiResta: Yeah, so I studied disinformation campaigns. I got involved in this looking at American conspiracy theorists looking at the anti-vaccine movement and the impact that they were having on shaping conversations about legislation, so not their outreach to parents, but actually their outreach as a political force. As we began to have more and more disease outbreaks in 2015 there was that cause and effect on that front, but then there was also, as we tried to use legislation to deal with that, the way in which they were able to really galvanize a small group of people to have a disproportionate impact on the conversation by using things like automation, by doing very interesting strategic kind of social media coordination. Looking at that.

Tristian Harris: So, when you say automation, what kind of automation did they have?

Renée DiResta: The time it was mostly primitive bots, but primitive bots was all you needed in 2015, because Twitter wasn't very sophisticated about what would make something appear in a hashtag.

Renée DiResta: And so, owning the share of voice around a conversation was much easier because you just really had to kind of own the hashtag. And so they would just have these automated accounts that would be pushing out content 24/7, so anytime you pulled up the hashtag SB277, which was the hashtag for the bill, what you would see was their content, was their point of view, and I did a network map of the clusters of people in this conversation with a data scientist named Gilad Lotan, and what we saw was this remarkable, highly centralized, deeply coordinated, on message, a collection of communities that were leveraging automation. And then we looked at the public health voices who are sort of the other side of this debate. They would kind of like occasionally tweet. There was no real message to be on even.

Renée DiResta: There were hashtags like the hashtag vaccines work, but anytime the pro-vaccine side, which was so much smaller and less coordinated, would try to create a hashtag, it would just get taken over by the opposition, because they would just add it to their list of things that were being pushed out.

Tristian Harris: And it would flood the channel on top you mean?

Renée DiResta: Exactly. Yeah. And so the idea that this was a conversation was wrong. And so one of the things that I did was look at this conversation, look at this hashtag and then actually go to legislators and say, here are examples in which this is not really indicative of the balance of people who hold these points of view. So when you're polling your constituents and they're telling you, 85% are telling you that they're in favor of this bill to two to remove personal belief exemptions, which is a way to just kind of opt your kids out of getting vaccines for school, 85% of your constituents are telling you they want you to revoke that, to close that loophole. But 99% of the social media conversation is saying the exact opposite.

Tristian Harris: And so that that was the ratio, right? You would see 99% in favor of the anti-vaccine movement.

Renée DiResta: It was overwhelmingly anti-vaccine on social platforms. I don't know that we ever sat there and quantified the percentage of all messages through the entirety of the hashtag, but one thing that we did see was we would see these instances where the top 20 accounts were sending out 65% of the content. There is a really strange distribution by which 10 participants in the hashtag were dominating the hashtag, and that's because there would be these accounts that would just be kind of on 24/7. So it was really interesting to see that divide, and even if you look at vaccination rates in California, you would still see that 85% or so we're still vaccinating their kids.

Renée DiResta: But if you were to look at the social media conversation, it seemed like nobody was anymore. It was all done. So it was a really kind of profound to have a first person experience of that as a parent.

Tristian Harris: As a parent, right.

Renée DiResta: As a parent, as a person who was fighting to get that bill passed. I was in no way neutral on this whatsoever, just to be clear. I was really deeply surprised as we started to dig into how this conversation was taking shape, I am not saying in any way that these were Russians or that these were fake, that this was a point of view that wasn't real. This is a point of view that is very real.

Tristian Harris: Very real, right.

Renée DiResta: But the proportional representation of that point of view in conversation...

Tristian Harris: The amplification was not real.

Renée DiResta: The amplification was not real.

Aza Raskin: I hear this is a kind of consensus hacking. That if you can control what people hear, it starts to be like, "Well, I guess everyone else believes this."

Renée DiResta: Absolutely, and that's where, we didn't really have a term for it. I think Sam Woolley, who's a disinformation researcher came out with the term manufactured consensus, maybe six months after or something, Because I wasn't the only one who was looking at this stuff. There were researchers who were starting to say like, "Something is really weird here and we need to have a better understanding of how this dynamic, this online dynamic is changing our offline dynamics by influencing policy."

Renée DiResta: And I started writing these articles saying, what was, one of them was titled something really blunt, like Social Media Algorithms Are Amplifying Conspiracy Theories. And I couldn't...

Tristian Harris: And this is back in 2015?

Renée DiResta: I didn't have I, well I did... This was early 2016 I think was when that article came out. That was in Fast Company, and that was because we didn't have a terminology for it, but the same thing that we had seen with the anti-vaxxers in the U.S., all of a sudden there was... Do you remember the Zika outbreak?

Renée DiResta: There were these insane conspiracy theories going wild on Twitter that Zika was, a government created disease. The proliferation of these pseudoscience conspiracy theories and the way in which they were hitting people's Facebook feeds, people's Twitter feeds, because this was not yet seen as something that the platforms should have to deal with.

Tristian Harris: Right? If we take their model, it's like, if I'd ask you Renée, so isn't the solution to bad speech just more speech? Everyone has a right to say what they want to say. So if people are saying things that are crazy, just make sure there's more speech, which shouldn't that be adequate?

Renée DiResta: Well that was the, that was the state of the conversation in 2016 for sure. And there is an article that I actually co-wrote with Sam Woolley and I think Ben Nemo and two or three other researchers at the time.

Renée DiResta: It was in Motherboard, and actually one of the things that we say in there, which maybe is going to sound shocking now, this was related to ISIS and the terrorist Twitter bots was, maybe we should just be running our own bots. Because there was the idea that as long as the platform wasn't going to do anything about it...

Tristian Harris: Fighting fire with fire?

Renée DiResta: ... and that was the state of affairs at the time. Yeah, there were a lot of people who were saying maybe the solution to the ISIS twitter bot...

Tristian Harris: So it's essentially the solution to bad automated speech is more bad automated speech.

Renée DiResta: ... is more bad automated speech. And it's... and I'm almost hesitant to say that now because it sounds so terrible, but that was... There was a sense that either the platforms would have to come in and somehow change... restore some kind of...

Tristian Harris: Balance?

Renée DiResta: Balance. Nobody even knew what to call it, we were so lacking in vocabulary for any of this stuff then. We began having these convenings of researchers, platforms would participate, a lot of people in the conversation. And there was a deep recognition across all parties that we did not want the platforms to be the arbiters of truth. Now that was their term. And they used it quite often. We don't want to be the arbiters of truth, but there's a lot of...

Aza Raskin: Which of course presupposes that they aren't already the arbiters of truth.

Renée DiResta: Well, it's interesting because there's the content, right? And then there's the distribution pattern and if you kind of divorce those two things, the platforms didn't want to be seen as being disproportionately biased against a point of view or a piece of content. And that gets into realms of things like censorship and who decides what narrative can be said. But when you look at the problem from a standpoint of distribution, then you can say, okay, having accounts that are on 24/7 that exist solely to shift the share of voice, maybe that's not indicative of the most authentic view of what a conversation would look like.

Renée DiResta: So it became more of a conversation about integrity and how could we think about ways to ensure that we didn't get into the morass of is this point of view better than that point of view, or is this an opinion that is allowed to be said but not that one? But instead look at things like is there, what came to be called, “coordinated inauthentic activity” in which distribution is being gamed, and that is much more quantifiable. It's easier to detect. You can, you can look at it as like an anomaly detection point of view. This is not moving the way we would expect a normal pattern of orality to look like. Here is this account that's never been seen before that has 500,000 retweets on its first tweet.

Renée DiResta: How did that happen? Who even saw that first tweet? So you can look at it more as a signals and patterns phenomenon. And so, the content of the tweet, or the content of the post is not what...

Aza Raskin: You're not looking at what it is, you're looking at how it moves.

Renée DiResta: Exactly.

Aza Raskin: Sort of saying like, "Okay, well we don't know how a virus works or what its DNA is, but we can know that it's a virus based on how it's spreading from person to person."

Renée DiResta: That's the current kind of best practices as we think about how do we balance that right to expression with the recognition that there are certain distribution tactics that game the algorithm.

Tristian Harris: So you have this line from your early work on this. I remember when we were briefing Senator Mark Warner that if you can make it trend, you can make it true. Why is that true from a brain perspective?

Renée DiResta: So there's a lot of studies about repetition and people who hear a thing repeated over and over and over again. This is how manufactured consensus works, actually. You begin to believe that this is a dominant point of view or a thing that is true.

Aza Raskin: The illusory truth effect, I think is the, the cognitive bias here.

Renée DiResta: Yes, there is also the phenomenon that the correction never spreads as far as the original sensational moment or meme or whatever. And so that's where you get at things like something that sounds salacious, outrageous, sensational, is going to spread like wildfire because people want to share it. They want other people to know. It's not, it's not being done out of maliciousness. It's being done out of a desire to communicate information to your community. And the tools that we have built, they offer a velocity and virality or something can spread fast.

Tristan Harris: It's velocity as a service.

Renée DiResta: Yes. So it can spread fast and it can go far. And the way it hops from person to person and community to community is easier than it's ever been before. But it's very rare to see that very boring, usually quite mundane or scientific, or measured correction put out after the fact, achieve any kind of similar distribution.

Tristian Harris: Let alone stick. I think there's these studies where if you issue a correction, people actually do receive it, they nod their heads and then if you test them five weeks later, they completely go back to the false thing, which was more sticky in their brains.

Renée DiResta: Yeah. And so the stickiness and the repetition and the exposure...

Aza Raskin: I think it's the ease with which people can encode or remember a message is how true we think it is. So if you get hit again and again with a message you've seen a lot. So it's easy to encode. But humans also have a rhyming bias. So if something rhymes, you view it as more true.

Tristian Harris: Alliteration effects.

Aza Raskin: Alliteration bias, we view it if it alliterates who view it as more true. Confirmation bias than actually make sense in this frame because you're like, "Oh yeah, it's easily encoded because I just, I have all the bits of fits with my worldview so I'm going to view it as more true."

Tristian Harris: And this is where we realized that we're sort of always fighting fire with fire, because it's really just understanding our own nature. We have this line together. Freedom of speech is not the same as freedom of reach. Now why would we encode a line like that? And it's like, well, it's alliterative and it's more sticky and we're trying to put it out there because we're saying that we have to have a new conversation from free speech to talking about reach and amplification, but we're using the same techniques. So are we being... and it's the that you can't escape this arms race, this land grab for these sort of tuning frequencies of the human nervous system. You know, if that's a thing that resonates at that part of our nervous system, it becomes this race to who's going to play more chords, right on the human piano?

Aza Raskin: Hey listeners, we're going to pause our interview with Renée for a few minutes to jump off that last point. If these platforms weren't hacking into our nervous systems, what could they be doing? I know oftentimes we can sound anti-technology. We're not. The point is technology is shaping exponentially more of the human experience. So, what is the world we actually want to build?

Tristan Harris: I want to credit that Facebook has made a lot of progress in cleaning up the hatefulness and division of news feeds and the outrage-ification of the news feeds over the last two years. They've done a lot on that. The problem is they've probably only done that in English or a handful of Western languages, from which they're getting the most negative press coverage and government pressure. And all the countries, Kenya, Nigeria, Cameroon, Angola, South Africa, all these places where it's bad and there's probably very few people reporting on it or not getting nearly enough pressure. I mean, they're not building language classifiers that anti-maximize outrage in those countries. And you know, I think that's just a huge issue that has to do with their business model.

Tristan Harris: We can not have a world where what Facebook is about ... A lot of people think, "What are Tristan and Aza and Center for Humane Technology advocating for?" You know, just better ranking functions. What do you want us to do, make people read the New York Times? Should we give people fewer notifications? It's like, no, let's just change the purpose of what Facebook is for. So long as its primary interface is peer-to-peer sharing of links and content, it is vulnerable to the problems of a race to the bottom for attention, but it doesn't have to be for that.

Tristan Harris: And even Mark Zuckerberg himself, in 2005, when he described what Facebook was at the very beginning, a year into Facebook, he didn't say it's for getting content out to people, and publishers, and making sure that people can browse the content they love, or something like that. He said it's a social utility that's like an address book for keeping track of and connecting with your friends. And it could be about bringing us back to our local environments, embodied environments where more time at dinner table conversations with friends, more time having rich discussions, more time, doing things we love, knitting groups, sewing groups, church groups, reading groups.

Tristan Harris: That fits some of the rhetoric that they're making about Facebook groups, but I imagine that the new Facebook groups are only embodied groups, meaning places that you gather with physical people. Because people tend to be calmer, especially when they meet with people from the other side, when they're in person. We're leveraging the effortlessness of our instincts capacity to find more common ground or more trust. People are more reasonable when we meet them in person, than when we see them from a distance. Podcasts, we get a more reasonable view into people when we hear them that way, versus when we see 255 characters of text on a screen.

Aza Raskin: This is, I think one of the simplest things that all of these tech social platforms could do, is they could start by asking us about our values, right? If Facebook asked me whether I cared about climate change, and I said yes, then there are a number of things that Facebook could do, whether it's reorganizing the news feed, whether it's about telling me about groups that I could go participate in right now, showing me all the people. Instead of it showing me the people that are getting engaged, it shows me all the people switching to a plant-based diet, which I've said I've wanted to do anyway. All of these things help my impulses line up with my values because they can ask, and not a single one of these platforms asks us about our values right now.

Tristan Harris: I mean, and this is the design research project, which is how will and how can software actually be in conversation with our values, as opposed to our lower level nervous systems? The part of this Copernican Revolution of moving the moral center of the Human Universe from the authority of human feelings, the triggers in her nervous systems and behaviors, and calling that our truest revealed preference, and say, "No, no, no. That's our revealed nervous system”, but are true revealed values and moving to a revealed values model has to involve a new design, that's actually good at eliciting what about this is important to you?

Tristan Harris: It could be as simple as, let's say you post an article about climate change. When I do that, why is that meaningful to you? People have to be articulate about this, but I think this is the research project. This is what all of Silicon Valley needs to engage in, is what is important to us? And right now software doesn't ... We don't even know. That's an open research question. It's exciting. Let's figure out how to get into people's values, instead of get into their nervous systems.

Tristan Harris: All right, let's go back to the interview with Renée.

Renée DiResta: The memetics is interesting because ... Memes is not just cat pictures, right? But memes is units of cultural transmission, ways in which we encode meaning. Memes is using the Dawkins Sense, genes of culture, right? Cultural genes. So, the building blocks, the foundational building blocks of culture, which spread from person to person. And there's that stickiness in memetics, which is why the thing that I think about a lot, is ways in which platform design choices have facilitated memes as the dominant form of information transmission, right? So, you want a square picture, you've got to communicate your information in that picture. People aren't reading long form articles quite the same way, so it's a way to take something. There's a visual. You remember the visual, you remember the alliterative message. Usually, it's quite simple. You're going to get max two sentences in there, maybe even just one. It's a fundamentally different way of transmitting these short bite size, completely nuance lacking pieces of information very rapidly, and they lend themselves to this virality.

Tristan Harris: Well, let's jump into that because this comes up so often. People say, "Well, hold on a second, Renée. You, Tristan, and Aza, hold your horses here. We've always had propaganda, we've always had advertising. Russia has always been doing this. We've always had fake news. It goes back thousands and thousands of years. Aren't you all just overreacting about this current state of play in 2019 with how platforms work?" What's your response to that Renée? And Aza, feel free to jump in too.

Renée DiResta: Yeah. I say, well absolutely propaganda has always been around, right? Propaganda is information with an agenda. There will always be information with an agenda because this is how we persuade people to points of view. This is how we get people elected. This is not new.

Tristan Harris: But then there's no cause for alarm. We're in exactly the same state as we've always been.

Renée DiResta: I talk a lot about the unification of three factors, which just came about as social platforms evolved, and that's the mass consolidation of audiences on to five places, which means that you no longer have to reach people in every local paper, or local radio station, or whatever. You have this deep consolidation on to five platforms.

Tristan Harris: So, I have to go to five places, instead of going to 100? It's much cheaper is what you're saying?

Aza Raskin: It's cheaper.

Renée DiResta: Right. There's more focus. You can direct your energy and reach millions of people in communities online. There's target ability, so these platforms offer the ability to reach people. That's how they make their money, which means that if you want to reach a particular group of people, you can, in a way that you never could before, so you have that granularity where you're able to reach people, not just according to where they live or what they read, but who they are, which is a very different thing.

Renée DiResta: And then, the third piece is the gameable algorithms. And that's where you get at the unintentional lift that you get from the platforms. And so you see this ... when you look at, going back to 2016, before Twitter decided to take action against the manipulation mobs, whether Russia, or domestic ideologues, or spammers, you would see nonsense trend regularly because they just knew that if they could get it trending ...

Tristan Harris: They could make it true?

Renée DiResta: Well, but not only that, they could also get it into mainstream newspapers because journalists were on Twitter. So, you could get extraordinary distribution there.

Tristan Harris: Let's talk about that too because you brought up this really important point when we first started having some conversations together, that once you make it trend, the reason you make it true ... Conspiracy, for example, if the media reports on it, then they make it true because they're spreading it everywhere. If the media doesn't report on it, then it's a media conspiracy that they're not reporting on something that's true.

Aza Raskin: Double blind.

Tristan Harris: It's a double blind. I've tied your hands behind your back. Magicians do this all the time. You know? You create a false choice. Do you want to talk a little bit about that?

Renée DiResta: What we kept seeing was something would trend ... And if you remember, Facebook had a trending topics thing too, and this was a huge deal because there was the ...

Tristan Harris: That's right. People forget about this because it's no longer there.

Renée DiResta: Yeah, because it happened more than six months ago.

Tristan Harris: Only a year ago. Exactly. You mean, yesterday.

Renée DiResta: Six hours ago.

Aza Raskin: Downgrading human attention spans, that they can't remember six months ago. Go on.

Renée DiResta: With Facebook trending topics, there was this controversy, came to be called Conservative Gate, where conservatives felt that Facebook was somehow silencing or preventing conservative topics from trending. And Facebook's response was to eliminate human editorial curation from trending topics entirely, just to avoid any potential appearance that human bias was setting the agenda for what people were seeing in trending.

Renée DiResta: And immediately after that happened, absolute nonsense started trending with regularity. There was an article about Megyn Kelly getting fired by Fox News lately, blatantly false. There were articles by conspiratorial sites that nobody would think of as in any way reputable or that would make their way to the top. One night I logged in, and there was like a witch blog talking about a new planet that was trending. Literally, a witch blog, I'm not kidding. I took a screenshot and I sent it to a data scientist at Facebook, and I said, "This is a disaster."

Tristan Harris: And this is because, just to slow it down, people are hitting share on these stories.

Renée DiResta: Share, share, share, share, share.

Tristan Harris: Share, share, share, share, share.

Renée DiResta: And that's my assumption.

Tristan Harris: And there was a one-click instant share at the time, is that correct?

Renée DiResta: Well, Facebook, I think that there was one-click instant share at the time. And then, there was also ... Trends were moderately personalized, I imagine. And then, there was also that subset where you could click into the science trends. There was a science trend one day about how dinosaurs didn't really exist. These were things where I was just like, "Oh my goodness. This is just all being gamed." They don't want to be seen as putting their fingers on the scale, and so anybody who comes in with a click farm-

Tristan Harris: But aren't we just giving people what they want, Renée? I mean, isn't this just a ... If you have two billion people jacked into a trending topics list of 10 things ... And I think the thing people miss about this, attention is finite.

Renée DiResta: Yeah.

Tristan Harris: When these things start flooding the channels, it's not as if there's this infinite supply of the alternative ecosystem. This garbage starts to fill up the airwaves and becomes the airwaves. It becomes the new normal. When manipulation works at the first layer, you're trying to do something directly. You're pulling directly on the puppet strings. But once you kind of get into them, let's say you implant a habit or you implant a deep seeded belief, you don't have to then pull in the puppet strings anymore. You can take your hands off and watch the puppet go walk around in the world, shouting these beliefs that are now running through their mind, and they're automatically pursuing that agenda. And this is happening at many different scales.

Renée DiResta: A lot of what we saw with Russia, was the building up of tribes. They do that, not by making you hate other people, but by making you have very strong points of view about your identity and your identity as a member of this group. And so-

Tristan Harris: So what's an example?

Renée DiResta: For black women, for the content targeting black women, a lot of it was just focused on family, what it means to be a black woman in America. Inspirational images of aspirational black marriages. Black Fatherhood was a really big theme. Some of that was-

Tristan Harris: And these are Russian trolls pushing memes on black fatherhood.

Renée DiResta: Yeah, yeah, yeah. Absolutely. Yeah, there's black hair, a lot of beauty, fashion. Black don't crack, which was a sort of phrase that ... I would reach out to black women researchers who studied trolls. Also, my dataset was NDA'd, so I could not send them the memes, but I would say, "So, what do you know about this hashtag?" Because this is not content that I am regularly pushed as a white woman. So, I wanted it to get some sense of, is this something that they made up or is this something that they appropriated from actual black culture pages?

Renée DiResta: And what we would see, as we got more and more into the data set ... This was a six month study or so, eight month study, ways in which they were just taking and repurposing hashtags, and phrases, and memes and visuals, including that real black women had posted themselves. So, this is my Tumblr account with my picture of me, and then they would take that, and they would share it, and they would say, "Look at this queen." Or something.

Renée DiResta: One of the reasons the dataset is not public actually, is because there are so many images of people, and they're real people.

Tristan Harris: Real people.

Renée DiResta: And that's because they would put their pictures on Tumblr, or they would put up their photos, and then that content was seen as something that could be appropriated. And so the-

Tristan Harris: By the Russian trolls?

Renée DiResta: Yeah, by the internet research agencies. The pages that were targeting black women, the pages that were targeting ... This wasn't just ... Of course, it wasn't just the black community that was the recipient of this. The black community was the majority I would say.

Tristan Harris: Right.

Renée DiResta: Most of the content really did, they leaned very hard into the black community.

Tristan Harris: Right. I noticed we're not saying something about, "Oh, they got duped," or something like that.

Renée DiResta: No, not at all!

Tristan Harris: It's not this at all.

Renée DiResta: No, no, no.

Tristan Harris: It's actually just they're playing to pride and identity. You would never know.

Renée DiResta: Yes.

Tristan Harris: I mean, we see images all the time of different kinds of pride, but they ... So, why did they go after-

Renée DiResta: And they did the same thing with southerners actually.

Tristan Harris: Uh-huh.

Renée DiResta: The narratives about the confederacy were not ... They were not rooted in hate. It was like, "We are proud descendants of this group of people who fought this war, and this is our flag." And so, it was very much a rally around that pride. Very rarely was it positioned in opposition. They began to position it strongly in opposition when the confederate monuments were coming down. And then even then, that framing was about your identity. This is an attack on your identity. This is isn't a front to you as a southerner. They had a nuanced view of how the right operated too, in the sense that pages targeting older people leaned more into narratives of security, Ronald Reagan, lots of images of flags.

Tristan Harris: Just imagine the history of Russia pushing Ronald Reagan memes.

Renée DiResta: I know. There's irony there.

Renée DiResta: The younger leaning Russian stuff was much more of the snarky ... There was a meme that was are you team conservative or team cuckservative?

Tristan Harris: This is younger conservative you're saying?

Renée DiResta: Yeah. So, it would lean more into the ... The had pages targeting the tea party. They also had pages targeting more like the pro Trump right. So, they did have that segmentation, and they would decide how to ... They wanted to erode support for institutional Republicans as well, so there was a ton that was anti-McCain, ton that was anti-Lindsey Graham, particularly when Lindsey Graham was at loggerheads with President Trump or then candidate Trump. There were anti-Ted Cruz , anti-Marco Rubio. During the primaries when they wanted to kind of bolster support for then candidate Trump, on the left the political content took the form of anti-Hillary. Bunch of stuff that was pro-Jill Stein.

Renée DiResta: When Bernie was still in the ring, pro-Bernie Sanders. When Bernie was no longer in the ring, the conversations about the ways in which the Democratic Party had wounded Bernie voters. This is all rooted in real grievances.

Tristan Harris: Right.

Renée DiResta: There is some truth to a lot of this, and that's what makes it insidious because the hardest thing to respond to is always, yes, but we hold this point of view. Who cares if the Russian said it? Because we hold this point of view also.

Renée DiResta: You're deepening a sense of an identity that someone already holds, right? When you're looking at that original ad targeting, if they're targeting an ad for a Christian page, they're targeting it to people who are receptive to that point of view already because they're Christians, and that is a perfectly normal thing to be-

Tristan Harris: Right ...

Renée DiResta: And that is a perfectly normal thing to be.

Tristan Harris: Right. Nike might want to also target Christians.

Renée DiResta: And if you are a Christian you should want to find Christian content. Exactly.

Tristan Harris: Or whatever.

Renée DiResta: And the same thing with if you are a Muslim. That same thing, if you are a Southerner and you want to find your Texas Pride page.

Renée DiResta: I mean, I'm a New Yorker. I have New York pride, right?

Tristan Harris: Right.

Renée DiResta: So there's no... That sense of who you are, your identity, your place in the world. You find people who share that identity. You deepen that affiliation and then you add on the propaganda layer.

Tristan Harris: Yeah.

Renée DiResta: Then you add on the call to actions.

Aza Raskin: This is a really strange idea. Russia is creating propaganda for us that reinforces our world views. We want to stop one more time here and really get into this question. If Russia is just giving us things that we want, things we already identify with, things that make us laugh in agreement, then what's the problem with that? Tristan and I talk it over.

Tristan Harris: Well I think the question is, when the receiver of that persuasion is unaware of the source of it or the motives behind it how do they feel? Like, I love getting that deal that says $100 off buying this thing. I'm like, "Oh, that's amazing. I'll take that." But I don't really know the economics or who is paying for it or why they want that to happen. Is there something is going to happen afterwards? Or maybe I'm signed up for something or I just gave away some piece of information about be or now the voodoo doll about me is way more accurate, and I didn't know how and that's going to cause me trouble like five years from now.

Tristan Harris: I mean, let's take this example that just happened. There's this app called... I mean, it's so funny. People say, "Oh persuasion. What a conspiracy theory. Isn't Tristan and Aza, they're just exaggerating this whole thing. This isn't really real." Just yesterday, this app called FaceApp I think has been downloading 150,000,000 times, and all it does is you take a photo of your face. I think you do like a 3D scan type thing and then it does a deep fake style, almost like CSS style shape over your face to make you look older, but it's really accurate. So it really makes you look, like this is what you'll look like when you're older. And it plays in to just the core, it's like the perfect persuasive cocktail like mix, mixer shake up of persuasive ingredients.

Tristan Harris: So it's vanity. You're the star of the show. It's about you. It's about what you look like. People love that.

Aza Raskin: Yeah.

Tristan Harris: What am I going to look like when I'm older? Two, social validation. So what do all of you think of what my old face looks like? Like what do you think of that? Isn't it kind of funny? Don't I look actually kind of attractive when I'm older? And then the third thing, which is social proof. Hey, everybody else is doing it. It must be okay. And guess where this app was built. Russia.

Aza Raskin: Was it really?

Tristan Harris: I think the company is called Wireless Labs.

Aza Raskin: Wow.

Tristan Harris: Yeah, so there's this thing. And people say, "Oh man. All those dumb gullible people over there that got influenced by Russian propaganda. Wow, how vulnerable they are. Like good thing I, the smart one over here, I would never be influenced by that." I have like at least 100 friends in my newsfeed on Facebook who have actually installed this thing, and I have a lot of smart friends. I mean, it's not correlated to intelligence, right? Or even critical thinking, right? It's about some core nature.

Tristan Harris: Aza, you and I were at this dinner and someone we know who is a CEO of a major tech company who knew Zuckerberg in the early days, said that Zuckerberg said this fascinating thing, "That every human being responds to social validation." Not one human being does not respond to social validation. It is a universal, and if you own that, you own that fulcrum of what motives people on a daily basis. That's why likes are so powerful. That's why having your profile photo different tomorrow and having that visible for other people to see or respond to is so powerful. That's why FaceApp is so powerful.

Tristan Harris: But when Russia just basically got the names and photographs, close up photographs of 150,000,000 Americans for the 2020 elections if they want to use it. Now I'm not saying this was done by the FSB. Although, it definitely could have been. Who knows. But certainly because this app was built in Russia, it would not be very hard for Russia to commandeer that database and say, "Great, what do we want to do now? And how about all that deep fake stuff we can do. We'll start calculating who people's friends are. We'll make up posts. We'll like..." There's a whole of bunch of stuff you can do now that you have that dataset. And people think, "Oh, only those gullible people could be influenced by propaganda."

Aza Raskin: No. Well the thing is in Kenya, Nigeria, and South Africa it was a third of respondents in this one Nieman lab study, said that they themselves had spread false news. So we see that the effect is like at the scale of a third of a population, and that's just the people that they surveyed. It makes me think of-

Tristan Harris: Who admitted it though.

Aza Raskin: Correct.

Tristan Harris: They said that they knew that they had done it.

Aza Raskin: Right. And who admitted it, which means that the number is almost certainly higher because who wants to admit that?

Tristan Harris: Oh, like at least double. Yeah.

Aza Raskin: Yeah, exactly. And we always get this question. Renée gets this question all the time of, "Okay, but did this stuff actually influence elections?" And of course, that's a very hard thing to know in part because every single human being is in the control or rather is not in the control group. They're in the experiment, so how do you? Well it's not like there's a second Earth in which Facebook does not exist and is not going these things.

Renée DiResta: You know, I get asked the question, did it swing the election? And the answer is I have no idea, because we just didn't have that granularity. I can tell you that 500,000 people followed a particular page. They were likely of a certain demographic, but I don't know where they lived or what their prior position was or anything. It wasn't part of the data that I had.

Renée DiResta: But one thing that was very interesting, was in the week leading up to the election the content targeting the right was all about anger. It was phenomenally angry. It was we need to get ready to have an armed insurrection if Hillary steals the election. It was we have to vote to stick it to the elites. It was constant anger. It was just constant anger to drive people to go take an action. Go vote, go vote, go vote. Not even, go vote because we love President Trump so much. It was go vote because she can't win and if she wins it destroys America. So that was where you would see this. We have to be ready to riot.

Renée DiResta: Meanwhile, on the left and in the black community it was apathy. So it wasn't anger at all actually. It was just, why would we get out of bed for this? This isn't for us. It was an interesting-

Speaker 1: Didn't you have an example where they were posting photos of like cute black families or something like that during the election week? It was sort of-

Renée DiResta: It was still leaning in to... They posted a lot of inspirational stories about black youths in particular. That was always framed... Actually, I loved reading them. I thought it was great. Like as somebody who was reading these stories as I was going and saying, "Oh, that's an interesting story. I hadn't heard about that." But it was being framed like this was the narrative the media doesn't want you to see. So all sides got that, this is the narrative the media doesn't want you to see. So that erosion of trust in mainstream media. There were constantly memes about CNN. Who was controlling CNN. There were regularly posts about... And it wasn't just targeting the right, right? Because that's a pretty common narrative. It was the black community pages that they built all pretended that they were independent black media telling the stories that mainstream media wasn't telling. Now the irony-

Tristan Harris: Now why does that work?

Renée DiResta: Is that they were actually going and grabbing these stories from American media. Cutting and pasting them and repurposing, but that of course, you know. Because I started tracing these stories back, because I'm saying, "Okay, where are these stories coming from? Are they making them up or are these people real?" You know, these stories, these pictures. I've got pictures of people's faces, is this actually the thing. There was one about an African-American kid who ran a GoFundMe for a medical device that he wanted to build. Sorry, it was a Kickstarter. And they wrote about this particular story like three times as he-

Tristan Harris: And this was during the election week?

Renée DiResta: It was to keep kids from dying in hot cars. No, this was like a little bit earlier than that, but they kept returning to these stories, and each of their fake black media pages, of which there were about 30, would post the story on a different day and would repost the story again. So you would see them taking their content that was resonant and reusing it. So they knew what their wins were and like any good social media operator they would double down on those. It was interesting to see how they did it.

Renée DiResta: But a lot of the narrative leading in to the election week for the African-American community was very much focused on this isn't our country. So there were stories of police brutality incidents, which were a common theme throughout, because again, this is rooted in real grievances.

Tristan Harris: Right, these are all real grievances.

Renée DiResta: Right. But they had that and then the frame that the used for that was, so we shouldn't vote.

Renée DiResta: And so this... We're second-class citizens in this country. They treat us terribly, so we shouldn't vote. And so that was where you started to see a lot of the “they” language. A lot of the other. Why would we participate in this process that is not for us? So you build up these pride based groups. People who are proud because they have a deep connection to an identity and then you turn that on in an advantageous way when you want to be manipulative. And so it was not black people shouldn't vote, it was as black people we shouldn't vote. And it's a subtle inflection, but it builds on the community that you're in.

Tristan Harris: It makes it incredibly hard to say, "Let's turn that off." Because how can you? They're just saying things that people already feel in much the same way that people say Trump says things that people already feel when he's saying extreme things. What is the recourse against this? What do you do?

Renée DiResta: Well, this is where it gets really hard, right? None of the Russian content, or very little of it, would have come down on any term of service violation, because it wasn't really objectionable, because these are all positions that real people hold. And so we would get in to these interesting conversations, particularly when we'd talk about truth and not wanting to be arbiters of truth. And this was again, how the conversation came back to integrity. It's a really weird nuanced thing to have to work through, which is what is an authentic Texas secessionist? Who is an authentic Texas secessionist?

Tristan Harris: Right.

Renée DiResta: So there are people in America who are Texans who are secessionists.

Tristan Harris: Right.

Renée DiResta: And that is there sincerely held belief and under freedom of expression they have every right to express that sincerely held belief. How does Facebook decide if the Texas secessionist page is run by a quote, unquote authentic Texas secessionist. And this is where we get at some really challenging nuances where this is where you've got this collection of like... Platforms have access to metadata. This is where you see the changes Facebook has made where it tells you the regions page managers are from and stuff like that, but is there an expat Texas secessionist living abroad. You know what I mean?

Tristan Harris: Yeah.

Renée DiResta: It's just a mind-boggling collection of really hard. It's a very hard problem, so we do try to get back to dissemination patterns. We do try to get back to account authenticity. Is this account what it seems to be? Ultimately, propagandas have to reach an audience and so that's one of the things we look for, is when you have to reach mass numbers of people what do you do to do it? And what makes that action visible.

Aza Raskin: We're going to end part one of Renée's interview here on this key question. Senator Mark Warner said that for less than a cost of a F-35, Russia was able to influence our elections. This kind of viral propaganda is a new kind of thing. It's an autoimmune disease that uses our platforms against themselves and it uses our values against us. I thought of a metaphor as I listened to Renée's description of these memes. In April of this year there was a paper published in Nature of research at Harvard. They stuck probes in to the brains of macaque monkeys and then trained an AI to generate images who's goal was to get the monkeys' neurons to fire. The AI started with visual noise and it kept iterating and tweaking and dreaming up new images until the neurons were hyper stimulated, firing way more than they would for any natural image.

Aza Raskin: The images that resulted were glitchy and surreal. Sort of like a bad trip. You could see other monkeys in collars and faces with masks. And I thought, "Isn't this sort of like what we are doing to ourselves as a species?" Testing hundreds of millions of pieces of content against the human psyche. And what emerges are bizarre things that we might not understand, but that work. So how can we make this kind of viral propaganda first more visible, and then second, how do we keep it from going viral in the first place? How do we upgrade our immune system. Here's a clue from Renée.

Renée DiResta: This idea that we have of global public square. That's actually ludicrous. That should never have existed. The idea that shouldn't even make sense to people, right? We don't even have a national public square. There's no such thing. And there's something to be said for smaller scales of communication.

Aza Raskin: Tune in to the next episode of Your Undivided Attention to hear part two of our interview with Renée DiResta.

Aza Raskin: Did this interview give you ideas? Do you want to chime in? After each episode of the podcast, we are holding real time virtual conversations with members of our community to react and share solutions. You can find a link and information about the next one on our website, humanetech.com/podcast.

Aza Raskin: Your Undivided Attention is produced by the Center for Humane Technology. Our executive producer is Dan Kedmey. Our associate producer is Natalie Jones. Original music and sound design by Ryan and Hays Holladay. Henry Learner helped with back checking and a special thanks to Abby Hall, Brooke Clinton, Randy Fernando, Colleen Haikes, David Jay, and the whole Center for Humane Technology team for making this podcast possible.

Aza Raskin: And a very special thanks to our generous lead supporters at the Center for Humane Technology who make all of our work possible, including the Gerald

Schwartz & Heather Reisman Foundation, the Omidyar Network, the Patrick J. McGovern Foundation, Craig Newmark Philanthropies, the Knight Foundation, Evolve Foundation, and Ford Foundation, among many others.