Center for Humane Technology | Your Undivided Attention Podcast
Episode 14: Stranger Than Fiction

Claire Wardle: Whether it was the brushfires in Australia, whether it was the downing of the plane in Iran, whether it was about the impeachment crisis, whether it's Coronavirus almost every story now has an element of mis- or disinformation connected to it.

Aza Raskin: That's Claire Wardle, the co-founder and US director of First Draft, a nonprofit that trains journalists how to combat mis- and disinformation worldwide. Claire says the problems journalists are confronting overseas rarely get any press coverage here in the US.

Claire Wardle: Certainly, the conversations in America do not recognize what this looks like globally.

Tristan Harris: In fact, the conversation seems to be stuck in 2016. While we've been talking about the four-year-old threat of fake news, Claire's been watching whole new categories of threats go unacknowledged.

Claire Wardle: When people would use the phrase fake news, I would say, "Well, most of this stuff is not fake and most of this stuff isn't news."

Aza Raskin: If there's anything Claire can teach us, it's that most of these threats are so new we're at a loss for words.

Claire Wardle: How do we deal with a genuine photo that's three years old? It's a genuine photo, it's three years old, but the problem is the caption is placing it in a different context. Lots of research shows that audiences can't even consider that that photo would be three years out of date, because that's not a problem that they've ever encountered. They're not prepared to fight it because they don't know it's a thing.

Tristan Harris: Claire can describe this specific problem, she calls it false context and that's just one of the seven distinct harms that she's identified in her eye-opening presentation, The Seven Types of Mis- and Disinformation.

Aza Raskin: It's a presentation that takes us into the gray zones of information warfare, where bad actors and their unwitting victims slip between facts and falsehoods, news and gossip, sincerity and satire.

Tristan Harris: They can even share the truth and nothing but the truth, but still mislead users through the power of narrative and repetition.

Claire Wardle: Bad actors, they understand that they are pushing a particular frame. It could be about a illegal immigration. It could be about vaccine. Whatever it is it's about frames and narratives, and those of us on this side who are trying to push quality information are playing wack-a-mole with these atoms of content. I just think we're not going about this the right way at all. This idea like we're going to fact-check a thing, not how it works.

Aza Raskin: Today on the podcast, Claire Wardle tells us how dis- and misinformation really work.

Tristan Harris: Now keep in mind that we recorded this interview in late February before the Coronavirus appended life as we know it. Now that we've all retreated to our homes in this collective attempt to flatten the curve, this conversation about misinformation has never been more important. Because how we understand what the Coronavirus is and how dangerous it is and what we should do and how do we help each other is all mediated on screens.

Tristan Harris: All of us are living on screens, it is the new digital habitat. Everything we're about to explore with Claire, about what needs fixing in our information environment and how our minds really process that information are all the more important to tackle now.

Aza Raskin: One of the hopeful things here is that corona is like a tracer bullet, moving through our information ecology. It's a united threat, a united enemy that we can all finally face. So this is the perfect time for platforms to get it under control.

Tristan Harris: I'm Tristan Harris.

Aza Raskin: I'm Aza Raskin.

Tristan Harris: And this is Your Undivided Attention.

Claire Wardle: I actually have a PhD in Communication and I thought I was going to be a professor for the rest of my life. I was researching user generated content, so how did newsrooms verify content that came in from the audience over email? It was a very niche research topic and I thought nothing of it, and then the plane landed in the Hudson River in January 2009. The head of BBC Newsgathering called me and said, "Not one of our journalists in the newsroom knows what Twitter is, and somebody tweeted a picture of passengers on the wings. We didn't know how to find that picture, how to verify that picture or knew whether legally or ethically we could use the picture. Can you leave academia and help train all of our staff around the world on how to do this?" I called my mom and said, "I'm going to leave academia and that nice pension."

Claire Wardle: For the last 10 years I've been traveling the world training people. Then three years ago the question of, how do you verify information online became a thing that a lot of people cared about. What's interesting is 10 years ago I started my career teaching journalists how to find content and how it was going to open up their black book, how they'd be able to get different voices. All the things that we loved about this concept of what social media would bring. Now I'm training journalists to say, "Stop, be really careful, those sources probably aren't who they say they are." I mean in 10 years there's been 180 degree shift, and it's kind of astonishing.

Tristan Harris: What woke you up to this shift? You said, "Three years ago something shifted," what was that?

Claire Wardle: First Draft was founded in 2015 as a project of Google actually. Google recognized that the journalists were struggling to know how to verify particularly images and videos online. First Draft was founded with very little money to say, "Can you just build a

website to help journalists?" We taught journalists how to do geolocation on a video. How can you independently assess where something has been filmed? How can you do digital foot printing? How can you look at metadata to understand where a photo has been taken? We still use those same training materials now. Back then it was about, how can you make sure the material during a breaking news event is authentic?

Claire Wardle: Now it's, how do you know that, that trending topic is authentic? How do you know that this network of accounts pushing the same material is authentic? It's the shift again has been quick, but it's using the same tools.

Tristan Harris: When I first met you I think it was in 2017 right after the election. It was at MIT, the Media Labs Fake News Conference.

Claire Wardle: Oh yeah, the Mis Info Con.

Tristan Harris: One of the things I really appreciated in a world of simplicity and black and white thinking was your first desire to say, "Wait, hold on a second. How do we actually define an ontology, a framework for saying what's the difference between misinformation, disinformation?" These even new proliferating types of fake hashtags, fake trending topics, all these kinds of things. Do you want to just quickly take a moment and define some of these distinctions? The word fake news as we all know is really unhelpful and we want to have a dialogue about what the deeper stuff is.

Claire Wardle: Yeah, and if I'm being honest, this goes back to my academic roots. I mean I'm glad that I did my PhD, but it taught me to think about language and to think about terminology. I remember in my bedroom, it was kind of like November 2016 just after the election, I remember with Post-It Notes and been like, "Well here's an example that's this, that's this." I created a typology with Post-It Notes and at that conference I remember putting up the seven types of mis- and disinformation. It's like a testing ground and somebody tweeted it as people do at conferences.

Claire Wardle: Brian Stelter from CNN picked up the tweet and put it on reliable sources and since then this typology has taken on a life of its own. Worse I don't think it's perfect, it can definitely be built on. What it did was make people recognize that this isn't about truth and falsity.

Claire Wardle: The seven types start with satire, which interesting lots of people said, "Oh Claire, satire is a form of art, you can't include that." Well now we see satire used deliberately as a ploy to get around the fact-checker.

Tristan Harris: What's an example on that?

Claire Wardle: We saw this a couple of weeks ago with the Bloomberg clip, where he slowed down the debate which made it look like when he said, "I'm the only person on stage that has run a business." They basically went from candidate to candidate with reaction shots where they look stumped. They added chirping crickets as a kind of way of saying, "Look, nobody could answer my question." Then when people pushed back and said, "That is a

false video, that's disinformation," he quite rightly said, "It's not disinformation, I was trying to make a point and I was using satire as a technique."

Tristan Harris: Got it, so that was satire?

Claire Wardle: Yeah. Or we just see people label a piece as satire. It could be Islamophobic for example and when people push back they're like, "We're making fun of people who are Islamophobic." Satire is now something that we see as a tactic. Things like false connection, if you have a clickbait headline, the idea that you're taking somebody to a piece of content that doesn't deliver what it was promised. I argue that's a form of polluting the information environment.

Tristan Harris: You call that false connection?

Claire Wardle: Yeah, and then we talk about misleading content. We've seen this for years in tabloid newspapers, the using statistics in a way that's trying to slam something, bias through omission. All the techniques that we've seen around misleading content. We then talk about false context, which is genuine information but used out of context.

Claire Wardle: Say there's an earthquake tonight in Chile and I go and Google it and the first thing that comes up is an image of an earthquake but it was three years ago in Iran. I'm like, "Oh my goodness, what a picture," and I share that on Twitter. I'm like, "I can't believe this earthquake," it's genuine but it's used out of context.

Tristan Harris: It was true about a different thing and so the context has been collapsed?

Claire Wardle: Exactly. We're seeing a lot of the Coronavirus rumors are actually genuine photos of people with face masks from previous times. It's the easiest form but most effective, because as we know the most effective disinformation is that which has a kernel of truth to it. Why fabricate something when I've got something already that I can recaption? That's false context.

Claire Wardle: We then talk about imposter content, which is when people use logos or names. There may be a journalist that they trust, that name gets used to sell soap or is part of a propaganda campaign.

Tristan Harris: This is the Pope endorses Donald Trump?

Claire Wardle: Yeah, 100% but in that example you'd need the Vatican logo.

Tristan Harris: I see.

Claire Wardle: You'd need something that you know or you're led to believe that it's official.

Tristan Harris: Right.

Claire Wardle: Then we talk about manipulated content, which might be...

Tristan Harris: Would you also include an impersonated content like if someone starts a Twitter account called Tristan Harris one? It's looks almost like my Twitter account or the real Tristan Harris, right?

Claire Wardle: Yeah, because they're using your credibility.

Tristan Harris: Right.

Claire Wardle: Again, as we're scrolling through Twitter, it's very easy to create... I'd write Tristan with a one as opposed to an I, that's what I would do.

Tristan Harris: That's what people do with these like fishing a tag type things.

Claire Wardle: Absolutely.

Tristan Harris: Is the characters look almost identical but one and the I. When Donald Trump confirmed Gina Haspel who's the CIA director. The first response to his tweet was a fake account called Gina Haspel One or something like that. It was just, "Thank you Mr. President, I'm so excited for the job," but I think it was a Russian bot or something. The name was almost identical to Gina Haspel, but it just had one character off and people wouldn't know.

Claire Wardle: No, exactly. I'm sure we will talk later about deep fakes, but whenever I have a conversation with someone, I'm like, "Why I'm I going to spend the money and the time to do that when all I can do by putting a one as opposed to an I?" I mean there are too many ways to gain the system right now with very little work.

Claire Wardle: Then the last two, manipulated content, so imagine I have two genuine photos spliced together. We have an example actually that did very well just before the 2016 election of, it looks like people waiting to vote at a polling station. There is a guy being arrested by ICE and he's wearing his ICE jacket. We use it in training all the time because I often say, "Well, what do people think of this photo?" They look at it and they're like, "Um," and then after about 30 seconds they're like, "Oh I think it's false because everybody's looking down at their phone." I'm like, "Yes, because this is a genuine photo from earlier in the year during the primaries. Here is a genuine photo of somebody being arrested by ICE. You put the two of them together, with Photoshop it takes two seconds. If your worldview is believing that illegal immigrants are voting, why are you going to stop to make... Well of course that doesn't quite fit, all the shadows don't quite go in the same direction." Manipulated is taking something that's genuine and changing it or splicing it together. Again, much easier than starting from scratch.

Claire Wardle: Then the final bucket it fabricated content, so that's the 100% completely false, Pope endorses Donald Trump. Deep fake, something that's completely comes from nowhere. If I'm a bad actor, that's my least favorite of the seven buckets because I haven't tested it. I don't know whether people are going to believe that rumor. I'm going to have to spend money to fabricate a video, why I'm I going to do it? Of all those seven buckets or

types, the thing that we see the most of is this false context, which is genuine content used out of context.

Tristan Harris: Right, and then what is the difference in your official definition between misinformation and disinformation just so people have that?

Claire Wardle: Yeah, so disinformation is false information that is created and shared to cause harm. Misinformation is also false information but the people sharing it don't realize it's false. We see a lot of that during breaking news events, where people see something, they don't know it's false and they share it trying to be helpful but actually we have become weaponized. That if we didn't share, if we were more thoughtful and slowed down all the things that you talk about, we wouldn't have so much of a problem.

Claire Wardle: In this whole space, the number of bad actors who are really trying to cause harm is small. The problem is us and because the bad actors are very good at making us do this by taking advantage of our fears, emotions, all the stuff we're probably going to talk more about, then it really works.

Tristan Harris: One thing listeners should know is just, when I think of Claire I think of someone who is on an airplane flying between election to election, like swinging in with a cap. You have this like SWAT team that you bring in to try to prepare news organizations in different countries around certain information concerns.

Tristan Harris: What's an area of harm that you think is underappreciated? What are some of the alarmist concerns that have been over appreciated? Just to kind of calibrate, like the public debate about this topic it focuses on fake news and then the Russian bots. That's like, if you're drawing out a continent map you would think that's 80% or something of the problem. I don't know, just something like that, but then the actual map, if the map is not the territory, where should our concerns be?

Claire Wardle: This is a great question because we obviously work globally and I feel very luck to do that because when you spend any time in the US the focus is almost entirely about political disinformation on Facebook. The reality, the rest of the world its health and science misinformation on closed messaging apps. The Coronavirus has made people recognize that this is much more complex. It impacts so many different topics.

Tristan Harris: You said closed messaging apps, by which we mean WeChat, WhatsApp.

Claire Wardle: Yeah, LINE, KakaoTalk, Telegram and yes there are some differences country to country and slightly different technologies. The biggest learning is all of this is about human psychology. What works in Brazil works in America and it's all about tapping into human fears. It's about tapping into in-groups and out-groups and it doesn't matter whether it's Nigeria or it's in a country split between two religions or India.

Claire Wardle: To be honest it's on one hand it makes it easier to understand, but on the other hand it's kind of depressing because you realize that it's technology agnostic and it's about technology activating the worst.

Tristan Harris: Right, and so do you want to give people a little bit of a hint of some of the places you've been and how these things have showed up around the world?

Claire Wardle: After November 2016 as you can imagine, there was much more concern about this globally. Just after the election actually we held some partner meetings with newsrooms, two in the US and one in London actually with European partners. French journalists said, "We're about to go into an election. We're concerned based on what we've just seen that we're not ready in France." We worked with French partners to work on a collaboration with over 13 newsrooms who said, "We don't think we can do this alone, we want to work together."

Tristan Harris: This is the election that ultimately Macron had won?

Claire Wardle: Yeah, May 2017. In that we tested this new methodology which we called Cross Check. Our belief is that no newsroom should compete when it comes to disinformation, because really journalists have never had to deal with falsity. That stuff ended up on the cutting room floor. What now is happening is the audiences are saying, "Yeah, we also care about the truth but can you help us navigate what's false?"

Tristan Harris: I want people to understand this innovated strategy that you came up with which was, in an environment and where there is media that is highly polarized, where the public doesn't trust different newspapers. Now let's say there's this new false information story that comes down the pike. If one of the, let's say the CNN of France says, "That's not true," if CNN of France doesn't have credibility, people aren't going to trust them shooting that story down.

Tristan Harris: The innovative approach that you came up with is, when you go into a country whether it's France or Brazil, what if we got all the news organizations together because people would trust it if like 30 different newspapers all said that this isn't true.

Claire Wardle: Yeah, so we did this project in France and what we realized is that, journalists working together was this amazing moment where people were teaching one another skills et cetera, et cetera. We now have rolled that out in places like Nigeria and Brazil. We've worked with journalists in Australia and Myanmar.

Tristan Harris: That's an interesting in a moment to relate, it's one of our earlier conversations with Rachel Botsman on the erosion of trust in society. In a low trust world people don't trust those who are even providing the corrections. I sort of see ourselves as a global civilization running around trying to just grab the last tiny, little bits of trust that we have in our institutions. What does have authority to shoot down something that may not be true? I see you as finding this nonlinear effect, that if we got these 30 newspapers to shoot it down, that might do something.

Claire Wardle: Yeah, and I actually have to credit, I don't know if people know this but First Draft is run between myself and my sister. She's based in London.

Tristan Harris: I did not know that.

Claire Wardle: She came up with this crosscheck methodology and Jenni doesn't come from a newsroom background. She said to me one night, she was like, "I think what we should do is get newsrooms to crosscheck each other's work." I said, "Can I just introduce you to newsrooms? This is never going to happen. They are not going to collaborate, nice try." She was just like, "I don't care, we're in a moment of crisis, we have to do it." I think it's another reminder of innovation comes from not having that sense of it will never work.

Claire Wardle: The very smart thing in France is, in the same way as here where people don't trust the beltway or elites, in France it's the same thing. Lots of people in France say, "I don't trust the Parisienne media." We had a coalition of yes, Le Monde and France Von Catan what you would expect but we also had Strasbourg 89. We had local newsrooms who were also in the coalition. Not only did they get to work with Le Monde and some of the big players, but they were much closer to their audiences. As the same as the case here, people are still more trusting of local news because they're more likely to know the journalist. They feel like it's more connected to their lives.

Claire Wardle: When we did the evaluation of the project, surprise, surprise some people said, "You know what, didn't trust half the people in your coalition, but I do trust my newsroom and to be fair, I'm pretty right wing and I think the Parisienne media left wing. I love everything that Cross Check did, but I trusted that because I trusted my news outlet."

Claire Wardle: That was when we had this moment of, no one organization is trusted by everybody. Is there a way that we can think about coalitions that might help audiences navigate this and recognize, well, maybe if we've got 10 different logos, then maybe there is something to this?

Tristan Harris: People trust it more.

Claire Wardle: It's not easy and people sometimes label this as cabals and newsroom shouldn't work together and this is collusion. Right now we need to try whatever we can and I think we did this in France, we did it in Brazil and had very similar feedback. There is something to this, whether we can do it in the US for me is, that's the big challenge for me in 2020.

Tristan Harris: The thing I think I find most interesting about this strategy is, this wouldn't appear to naked eyes like it would work because you have these newsrooms that are competing with each other.

Claire Wardle: Absolutely.

Tristan Harris: They're on opposite sides of the political spectrum and you said, "No, actually there's this kind of common good we need to protect here, which is the shared basis of truth and facts." Surprisingly they were willing to sign up for it.

Claire Wardle: Yeah, and I think this was because their newsrooms are very frightened about making mistakes. There were many newsrooms that said, "We don't necessarily have the skills in-house and we can't resource this and particularly at the smaller local level." That

competition piece is in, I can't do a French accent, but somebody said like, "There's no scoop in a debunk," like in an amazing French accent. What they were trying to say is, "Yes, we compete but we compete again around the good stuff about the investigations."

Tristan Harris: The exciting stuff.

Claire Wardle: "Not cleaning up a shit internet," I mean excuse my language.

Tristan Harris: Like, "We're not competing on brooms, we're competing on exciting, explosive material."

Claire Wardle: Yeah and to be honest, media has always had a pool system. Particularly in TV news, you don't send everybody with a camera to follow the president or the queen. One newsroom goes and then they share the footage. There was that belief which is on this, why we all wasting time chasing? Why we all verifying the same meme on Twitter when actually one person can do it? We can all look at the reporting and be like, "Yeah, we agree," and that's how it worked.

Tristan Harris: It actually worked that way, so there's a feed or something like that and then one news organization says, "This is the thing we think is a correction." The others can quickly validate it as opposed to everyone trying to research it from 20 different sides?

Claire Wardle: Yeah, because when you're checking the evidence, it's much quicker than starting from scratch. To be fair we saw many times when somebody was like, "Yeah, but actually we're not going to run it unless you actually get a quote from that person. Or we're still not 100% certain." It slowed down the process but it also meant there were absolutely no mistakes on any of the projects we've run now. When he talked to journalists afterwards, they would say, "It made me feel uncomfortable that I was forced to slow down, but surprise, surprise when I was forced to slow down it mean that the reporting was more accurate, was stronger and we didn't make mistakes." Like that is ultimately what we want.

Tristan Harris: I feel like this is something that as consumers of media and information we also have to gain a tolerance for. It's almost like sugar what's going on, because sugar just gives us that immediate hit. We like it but then we all know we would be better off if we just probably didn't have as much. We've been tasting this immediate access to, there's a breaking news story, Parkland shooting. I want to know in the next 30 seconds what the first report is of whether they know who the gunman is.

Claire Wardle: Yup.

Tristan Harris: Do we actually need to know that?

Claire Wardle: No.

Tristan Harris: How many and which human beings on planet earth when that happens, if you had to draw like a distribution curve, like how many people needed to know that within the first even let's say 24 hours? Did it actually consequentially affect our lives? I say this because I think we're in this uncomfortable tension, we have to trade some things. Like right now we say, "Well we want that immediate access to whether the Coronavirus killed exactly 57 people or in the next hour is it 58 people." I was checking the news this morning on Coronavirus, I'm very interested and concerned.

Tristan Harris: I guess it's just like, what's the humane clock rate in which information is dispensed? If we want the fire hose, we're going to live in this hyper noisy environment that may not be so bad if it's not consequential. When it's about whether or not you're going to go into quarantine for a month and lock yourself up with food or whether or not you're going to go inside because you're worried there's a, Las Vegas shooter, this just doesn't work. How do you see this tension resolving?

Claire Wardle: I mean I felt this on the evening of the Iowa Caucus, which was of course lots of mistakes happened. Seeing the media in that role was like, if there was no expectation that you would get the results for another 48 hours as is the case just with the Irish election it took three days, that's how long it takes. We also have to think about place of the economy here is, it's very easy to wag fingers at the media.

Claire Wardle: I mean right now the reason that people are so competitive and that every second counts is because they are desperate for clicks, because they're desperate for money and many newsrooms are struggling. What that means is, people are rushing and when a mistake happens, people are recognizing that the speed cannot, we have to slow down. We are being approached by more newsrooms to do more training because standards and ethics editors are like, "Yup, we cannot afford a mistake, not this year." I think it's not in anybody's interest to be quick.

Tristan Harris: Let's talk about the cost of mistakes, because I know something in cognitive science there's an effect. It's basically the first person to frame the debate wins because you set the initial frame. Let's say you hint that the shooter was actually this kind of disturbed military person, I don't know, I'm making it up. Now your mind is setting up an evidence accumulator, so your mind is pre-tuned to find and want to confirm and affirm evidence of that specific explanation. Which might be different than like, I don't know, something totally different happened. It had nothing to do with military, nothing to do with that kind of gun.

Tristan Harris: That other kind of evidence doesn't have the same or neutral acceptance by the mind, because the mind has been pre-framed by the first frame. When we think about the cost of misreporting those mistakes, it's like people don't trust the corrections. You kind of entrench yourself just not fully but in a deeper way in the first explanation.

Claire Wardle: No, and there's so much literature from social psychology about effective ways of debunking and issuing corrections but we know it's really difficult. We know even when you do it well, people are much less likely to share the correction. Even if you hear the correction, if you get asked two weeks later, you're more likely to be like, "Oh I can't

quite remember, but there was something, no smoke without fire." You tend to go to the original claim.

Tristan Harris: The original, exactly. This is the Briony Swire-Thompson's research at Northeastern University, I love it. It's just that people end up going back to the original belief, but so talk about corrections. You've learned a lot about in elections, what kind of corrections and what are the cognitive strategies for producing an effective correction in the human mind?

Claire Wardle: From doing the work that we've done, one concept that we've come up with is this idea of the tipping point. If you go searching for false information, you will always find it. If I go searching for some conspiracy Facebook groups with anti-vax content, I will find it. If I go looking on 4chan, I will find all sorts of things. Now it's very tempting going back to political economy, but if I want to headline that's going to get a ton of clicks, there's a ton of that stuff that I could write a piece about and I would get clicks. Of course, if you're a mainstream media outlet, you are giving oxygen to these rumors and conspiracies.

Claire Wardle: We talk about the tipping point to say, "Well if you report too early, you're giving oxygen to something that you shouldn't do. At the same token, if you wait too late to report on these rumors and conspiracies, it's almost impossible to bring it back." There's this sweet spot which from our work in these election projects, if you get it at the right moment and you get enough new dreams at the same time pushing out responsible headlines, we have seen evidence of slowing down the misinformation or having that misinformation taken down. That tipping point is something that's really crucial.

Claire Wardle: This is an example about France, but I think it's a powerful one which is, we saw a very sophisticated hoax website that looked identical to Le Soir, which is a Belgium newspaper. In fact, every hyperlink clicked back to Le Soir but the headline was saying that Macron was funded by Saudi Arabia.

Claire Wardle: Explosive content, we of course did very quick reporting and found out this was not true. Looked like there was some bots in Tehran that were pushing it, blah, blah, blah. We were like, "If we report on this it's irresponsible," but we sat on it and everybody was briefed, everybody had the reporting. We didn't do anything about it until Marine Le Pen's niece, who was the niece of Marine Le Pen who was running, she tweeted it. The tweet of that link meant that it suddenly passed the tipping point, and so then collectively Cross Check issued the report and we were able to slow it down and get that taken down.

Claire Wardle: Now, again, that's the perfect example I use in training, everybody loves it. The concept of that which is, how do you measure the tipping point? How are you sensible about that? There is evidence that you can slow this down, but that's part of our training with newsrooms to get them to think critically about when to report on disinformation. They should not be giving oxygen to everything.

Tristan Harris: It's like the thing they say about timing in comedy is like timing in corrections. You're essentially saying, "Hey look, we actually sat on this correction, we knew we had it but this wasn't the right time to do it." Then it suddenly spikes because of a natural organic event, which is Marine Le Pen's daughter or-

Claire Wardle: Niece, yeah.

Tristan Harris: ... niece posted and you jump on it.

Claire Wardle: A lot of the psychological theory also talks about the power familiarity in repetition. Again, when newsrooms are trying to be distinctive and they're competing, you don't have that familiarity and repetition. If you see a number of different outlets pushing out corrections, using similar language, similar imagery, which makes editors go, "That's not what we do Claire." Actually this is the new public service that is unfortunately required of newsrooms, which is to help audiences navigate our polluted information environment. In that environment we have to think differently and like familiarity and repetition work, and that's not what the news industry is about.

Tristan Harris: This reminds me of George Lakoff's work on if you're issuing a correction or if you see something false and there's different ways of reporting on it, put it into a truth sandwich. Truth sandwich is like two loaves and bread and something in the middle. You first say the truth, then you say the false thing that goes against that truth, then you repeat the truth at the end. If you just think about it in terms of quantities and repetition, you said the truth two times and you said the false thing one time. If you just add that up, you're fixing it.

Tristan Harris: One of our listeners in a past episode on hate speech said that, one solution, if there's a hate campaign that's later found out to also provide deterrence on future hate speech, is if you say to any poster of deeply hateful material, "If we discover your hateful content, we will later go back to everyone who saw the hateful content. We will back post twice as much positive content about the same minority group that you were posting about." We're just changing the saliency and repetition rate of the other positive story.

Claire Wardle: Yeah. I mean this goes to the core of a lot of training what we're doing right now with journalists, which is, how do you word headlines? In a headline with 40 characters, you don't have the chance to do that truth sandwich, so a lot of the evidence is where possible lead with the truth. Briony Swire-Thompson will say, "In the note graph you can talk about the falsehood because next to it you can talk about the truth." But the challenge is in the headline. We shouldn't really be repeating the falsehood in the headline. However, if I'm a journalist I'm going to say immediately, "Claire have you heard of SEO, Search Engine Optimization?" I have to repeat the rumor in the headline otherwise I won't get any traffic.

Claire Wardle: This is kind of a really fascinating question for search engines which is, I'm out there telling journalists, "Be really careful. Try not to repeat the rumor in the headline because you're actually reinforcing and giving more oxygen to the rumor. In an era where lots of

people just see the tweet, just read the headline and they don't read the nuance, we have to be really careful." Yet, I've got journalists who've said, "Claire, my newsroom has spent a fortune on search engine optimization training and we've been told we have to replicate the exact rumor in the headline to get the traffic."

Tristan Harris: The reason for that right is because for someone who's doing a search query about the false thing, they want to be found. Is there no setting where Google can say, I mean they have these meta tag names where you can say, "These are other search terms for this article and take them as seriously as these other ones." Only authorized journalism outlets can use these meta tags so that Google can privilege them in some kind of high authority type of way.

Claire Wardle: Yes, can you make that happen? I mean I'm not joking, this is happening every day in training rooms that we're doing.

Tristan Harris: Really?

Claire Wardle: We also talk to journalists about the difference between people searching versus stumbling upon and you're absolutely right. If there are rumors about Coronavirus, people are typing that into Google and we want that data void filled with something that responds to that. At the same time I don't want somebody stumbling on Twitter over a tweet that's repeating a rumor that I hadn't even heard of something that I should even be concerned about because our lizard brains remember the falsehood.

Claire Wardle: This is much more complex than these platforms that are designed to respond to, but I would love to have a conversation with Google. Which is like, how can we flag the fact that this piece of content should be connected to a rumor without basically repeating rumor everywhere-

Tristan Harris: The repeating rumor, yeah.

Claire Wardle: ... through it just to get the traffic?

Tristan Harris: You're talking to a set of listeners who are often in the tech industry or around in the surround sound of people in the tech industry. What do you really wish that they were doing or could do to help you more?

Claire Wardle: By working globally I also have huge sympathy with these companies who need global responses in order to scale the work that they're doing. I used to work for the UN and every year you had to change roles and move countries. I wish there was more of that in Silicon Valley because many of these companies tend to be in northern countries. They tend not to be lower and middle income countries and so many of the challenges here, whether they're linguistic, whether they're ethnic, whether they're religious, whether they're the different types of technology, whether they are the different conspiracies that have existed in these cultures for years, so much of that requires on the ground knowledge.

Tristan Harris: Could they even? I mean even with your work, so there you are, you've aggregated the 30 different newsrooms in France. The volume of things that are coming through the pipe, like trillions of content items are these matched? Is there just as many hoaxes and conspiracy theories as there are capable journalists waiting to pick up the phone to then shoot down the rumor? Or what kind of asymmetric situation are we talking about?

Claire Wardle: Yeah, let's take Brazil for this example, it's a huge country. Their news industry is struggling more than this one. They, almost every newsroom has a paywall. If I'm in Brazil here's my choices. I either can pay money to get access to a quality newspaper, well I can't really afford that. Or my WhatsApp and Facebook, because of the free basics means that I don't even have to use up data costs to access WhatsApp and Facebook. What I'm going to do? I'm going to go to WhatsApp where all of my friends are sharing screenshots of news sites.

Tristan Harris: This is something people don't know, I think our listeners and I'm not fully aware of. People are sharing screenshots of news sites because they'd actually have to pay money which they don't have the vast majority of people to look at news sites. This is because of the free basics program, which to quickly catch up those listeners who don't know, was a program that Facebook used to say, "Hey, you can get a cellphone and so long as Facebook and WhatsApp basically are the internet for your cellphone it comes with it on the cellphone, then you get the internet "for free"."

Tristan Harris: Then that privileges in terms of usage, the WhatsApp and Facebook as the internet, those are the primary surface areas through which people get their information.

Claire Wardle: Yeah, and the number of people who are sharing screenshots of news, let's just be honest are small. The number of people who are sharing memes and old images that are used out of context.

Tristan Harris: Oh false context visuals, yes.

Claire Wardle: Say for example, during the Brazil election we had a WhatsApp tip line, we received over 200,000 tips from the public about things they were seeing that they wanted help working out. The number one piece of content that was shared was a photo of a truck with what looked like a ballot box open in the back of the truck. The caption was, "These ballots have already been pre-filled in for Haddad." It was a genuine photo but the caption was false, that was not true, but it was shared everywhere. It's visual.

Tristan Harris: Do you know how widely it was shared?

Claire Wardle: I mean we received it over 1,500 times on our tip lines, so it was the number one piece of content. Again, lots of these countries you have lower literacy levels. You have people who have never had email addresses, the first time they've got their smartphones, all the stuff that we know we had to learn not to take scam emails from Nigeria seriously. We can laugh about it, but it's taken us 20 years to figure some of these stuff out.

Claire Wardle: We were actually at a conference in Singapore hosted by Google who had brought amazing people from Malaysia and Myanmar. They were these amazing people doing the same work that I do on a daily basis. I remember the first hour that was like seven minute lightening talks. By the end I was almost in tears because there was just incredible story after incredible story. I mean people saying, "I'm a mom but I do this as a fact-checker because I just really care. I see a lot of content now that makes it difficult for me to sleep." I want people in Silicon Valley to hear these stories.

Tristan Harris: What I think is interesting about this is that, from the outside, Facebook or Google or YouTube can say, "Look, we're hiring all these civil society groups. We're paying these fact-checkers, we're actually doing all the work with every single nonprofit on the ground in Myanmar, in the Philippines, in Cameroon. What do you want us to do? We're now working with all the groups that have those resources and have that local expertise." Then what ultimately that amounts to is conscripting them into a feed of essentially like the worst parts of society.

Tristan Harris: I think the biggest counter argument from those who are in the tech industries, like yeah, we know there's some bads, but there's also all these goods. There's some goods, there's some bads, who's to say? Or we think that basically those goods are enough to justify this. One way we could talk about whether the good balance sheet compared to the bad balance sheet is we could say, "Well how often are the goods happening and how often are the bads happening?" That's one way to do it. You could do it based on volume, like how much of the good is happening?

Tristan Harris: A different way to do it is on consequences, what are the consequences of the good things that are happening? What are the consequences of the bad things that are happening? If fake news, I think in Brazil it was something like 89% of the people who voted for Bolsonaro had belief in at least one of the top 10 fake news stories. They were complete crazy like out there fake news stories. If the consequences of the bad are authoritarianism rules the world because elections are debased in what people believe as the basis of their thinking.

Tristan Harris: When your brain is believing some basic set of cognitive frames and beliefs about the world and other people and the politicians you hate. Then on top of that your mind is looking for evidence to confirm what you already believe, if that's the cost of the bads, that's a highly consequential set of bads. I would argue that's a dark age entering set of bads, especially in these vulnerable countries where we just shut it down.

Tristan Harris: I sometimes just say, "Do we really need this? Is this really helping?" Or is there a safe way we just say, "Hey look, can we just do one on one messaging?" That's it, because anything more than that is just actually too damaging, it's too consequential.

Claire Wardle: Well, the good, bad debate is as you're saying, is way too simplistic. Actually what we're doing here is we're experimenting with people's lives in a way that we can't stop this snowball. I remember having a conversation with a Facebook engineer a year or so ago and I was saying, "I'm a social scientist by training, and what I worry about is we don't have longitudinal analysis." We've got psychological experiments right now mostly done

with students in large Midwestern American universities, deciding whether or not a corrections' policy works. If so, Facebook has flags, who doesn't have flags?

Claire Wardle: One of my hearers is Ifeoma Ozoma, she's a public policy manager at Pinterest. She said, "I don't want people searching for vaccine information on Pinterest. Until we know what the impact is, why should we have it on our platform and we're going to make that decision." My worry is, exactly to your point, when people are scared they are more likely to be supportive of authoritarian leaders because they're terrified. The strong man, this is a George Lakoff's stuff, that you want the strong father figure.

Claire Wardle: If you were terrified in Brazil about the fact that corruption has completely changed your country, you have less money in your pocket because you haven't dealt with the impact of the 2008 financial crisis, Bolsonaro looks like a pretty good deal in that situation. Same Duterte, you could argue the same with other leaders. What I think we don't understand is this drip, drip, drip, drip piece onto your point about consequences, we don't know. I'd rather that we stepped right back and that we tested some of these things without being like, "Oh God, in 15 years time we're going to say, "What the hell did we do?""

Tristan Harris: It doesn't mean shut off like-

Claire Wardle: No.

Tristan Harris: ... the entire internet or shut off www. like again, that's a different thing than let's create viral amplification of the fastest, least checked, most friendly to bad actor type of speech.

Claire Wardle: For example, Facebook live. I remember at the time knowing people at Facebook, who were kind of saying, "Hypothetically what if we created a tool that allowed people to just livestream from their kids birthday party, wouldn't that be great?" Of course, any journalist or foreign correspondent would be like, "I'll tell you what happened immediately, it's going to be terrorists, it's going to be suicide," and they went to a dark place. Facebook was like, "No, I think it's going to be about birthdays." It didn't take very long for Facebook live to really be de-ranked as a tool within the Facebook ecosystem because people realized it was too difficult to moderate and bad actors or for different reasons it was going to look bad.

Claire Wardle: That's how I feel now which is, can we look at all of the features, all of the tools? Pull back on the stuff that we know is potentially going to have more consequences and really go with the bits that we know and have been tested to really... I know this is very simplistic, but I think that there's so much experimentation. Things like more friction, which we know from all of the research is one of the best ways of slowing this stuff down. More heuristics, more labels, WhatsApp, give me more stuff to give me that context.

Claire Wardle: We know some of these stuff that works, and so what I would love to see is more of that less jazz hands about everything's going to be great. If Silicon Valley engineers could

spend more time in the pub with foreign correspondents and journalists, I think we'd be in a better place because you need some dark. Some people have experienced the darker sides of the world to say, "I'm sorry, like this is how your platforms going to be weaponized."

Tristan Harris: You're always laying so many things and I just want to double click on several of them. One was the fear, strong land base thing. I remember Brittany Kaiser who was on this podcast said that in Cambridge Analytica's psychological targeting that one thing they found was that, people who have the psychological characteristic of neuroticism always respond to fear.

Brittany Kaiser: It was only fear that really had a massive impact so they spent the rest of the super PAC's money on fear, yes.

Tristan Harris: We have a clickbait system, it's a lizard brain enhancing system and fear always works. It's a two-step process from when you're more fearful in general of Coronavirus, of what Russia could do to this country, of whatever it is. What a foreign power, you were going to go with the strong guy. If you just think about it that way, that you have this system that rewards fear over calm, truth, then it's sort of obvious to see why you would get kind of authoritarian people everywhere all at once.

Tristan Harris: To your earlier point, this is the largest unregulated psychological experiment done in history. Where's the IRB, review board? Where's the people who said who could be hurt by this experiment? I mean I remember back at Stanford, if you wanted to do a study with 10 people, you had to go through this incredible process to even run the experiment.

Claire Wardle: In those IRB processes, you have to go through extra steps if you're researching vulnerable people. I think about that in the global context which is, who are vulnerable? Well people who are newer to technology, people who have lower literacy levels. We know it and I think that's my frustration is that there hasn't been a recognition that we want to scale globally.

Claire Wardle: Yes, many countries have been transformed by this, and that's an important thing to remember, of course. I don't think there's also been that recognition of there are vulnerable communities here, communities that have been ripped apart or have religious tensions which mean, you've already got this like tinder box. That's how I feel, like there are many countries in this world who I would deem as a tinder box. For me this technology is the spark.

Tristan Harris: Right. What worries you about how governments are responding to this, especially the countries that are tinder boxes or are more vulnerable? What are governments doing? What have you seen that works? What do you wish they were doing?

Claire Wardle: I am very concerned. We've seen the passing of some pretty problematic regulation. Who writes this stuff? Politicians who are actually terrified that they will lose an election because of disinformation. They are not neutral actors here and I can understand why

there's been this panic about it globally and so they want to be seen to be doing something. Again, I'm a social scientist, we should not be doing any of this unless we have a foundation of empirical evidence and we have almost nothing.

Claire Wardle: If I was to say, "How much disinformation or misinformation as distinct categories is there? How is it different around the country? Are any of these solutions slowing anything down? Where's the benchmarks?" I can't tell you how much of this stuff is out there and what impact it's having on society and it's now 2020. Most people inside the platforms couldn't say this either and most social scientists can't say either because we don't have the data because it's inside the platforms.

Claire Wardle: In that context, no government should be passing any regulation because we don't know how much of it is out there and what impact it is having. What I would love to see governments do is hold the technology companies to account to say, "In order for us to have responsible regulation, you are going to need to work with us to allow us to audit what you're doing."

Claire Wardle: We did some work last year around auditing, what people saw when they searched for vaccines on Instagram, YouTube, Google and Facebook. We paid people in 12 countries to send us screenshots. We got over 500 screenshots, beautiful.

Tristan Harris: You did have a way to get that data yourself.

Claire Wardle: No other way to see.

Tristan Harris: You paid in 12 different countries people...

Claire Wardle: People to send us screenshots, that's an audit. Now that's what I would like to see governments do and I understand right now we see this tension between academic, saying to the platforms, "Give us all your data." Quite rightly the platforms are like, "No, we're not in a position to do that. We've seen a little bit with this social science world having to build differential privacy into our platform, it's hard. If I'm a government, I would argue that there is a way that data can be shared to simply say, "Show us, not the algorithm, but what's the output of some of your algorithms?"

Claire Wardle: Similarly, at the moment, we see different people. For example, Mark Zuckerberg talking about we've got new transparency measures. Now if I right now try and use that Facebook ad library API, it is an utter disaster. I cannot hold those ads accountable because I cannot monitor them because the API does not give me the information that allows me to do that. That's a fundamental problem.

Tristan Harris: Say more about that, what is the information that it's not giving?

Claire Wardle: It's buggy but also if I wanted to see what ads right now are running in Tallahassee and are they targeting people of color? Is there a vote of suppression campaign happening right now? I cannot do that, I can search by state, but I cannot search by those demographic categories beyond gender and age. I don't think that's good enough.

Tristan Harris: It's not legible in other words for research?

Claire Wardle: Yes, and so my concern is, these governments have been putting more pressure on the technology companies. We have had more transparency mechanisms, but actually Macron isn't testing the Facebook API. We have not good transparency measures yet they're allowed to tick a box and as I often say, they are marking their own homework right now. They write the transparency report to say, "We promise you we're doing better." It's not viable and so I want the government to work with the technology companies to get access to some of these data. To really have an ability to show whether or not these changes that are being promised are they making a difference.

Tristan Harris: That still puts this ad library approach, the ad transparency system still puts responsibility on society to monitor how much harm is there. The problem is like does society, does the nonprofit simple society groups have even the resources? First of all why did they send them for that job anyway? It's like why should we have to have an ecosystem? It's like we created all this work for people to just review how bad. It's like a gun manufacturer who's like, "We're going to provide reports every quarter on exactly how many people our guns have killed." We don't need transparency reports, we need systems that don't kill people. We need systems that don't target vulnerable populations.

Tristan Harris: Now the question's what does that look like and again, I think that in western markets things like the ad transparency API has some basic level of scrutiny from journalists and so on. It's better than not having it but that's assuming that there is this for the state that's monitoring the things that are going on in countries like Cameroon or Kenya or whatever. How many organizations are looking at the ad transparency API there?

Claire Wardle: Well they didn't have it. I mean Australia doesn't have the ad library.

Tristan Harris: Oh they don't, so it's just...

Claire Wardle: They went through an election. It's certain countries and you can imagine which countries have them.

Tristan Harris: Oh my gosh.

Claire Wardle: There are countries that have put more pressure on the tech platforms to say, "We require transparency." Australia had an election last year, we worked there, we've got a bureau in Sydney. We couldn't do the work that we were doing in the UK.

Tristan Harris: What was the cost of not having that transparency?

Claire Wardle: A complete black box. That's the problem. Of course, we don't have to have a long conversation about whether Facebook is correct to say, "We're not going to fact-check those ads." It is not correct to say that, "We have transparency methods that allow others to check whether or not." During the UK election for example, we found that one

of the parties, the conservatives, 88% of the ads they were running were using content that had been labeled as misleading by a fact-checker full fact, 88%.

Tristan Harris: 88%, yes.

Claire Wardle: That was because we had downloaded the content, worked with journalists and said, "This is problematic." Well so let's imagine 88% of ads running in Sydney based on misleading claims. Facebook isn't checking it and no Australian journalist can check it, it's not acceptable.

Tristan Harris: Those are the numbers, 88% in the UK and we know in Brazil 89% there, this does not look good. If we don't have the data we don't need more transparency that says how bad is the 90%, 95% rates in these other countries. We need to shut it down. What I'm curious about is, let's say five years from now, that's 2025 and we've transitioned to a humane technology future. We've reversed out of this heading towards the dark age where people don't know what's true. What in the period of that five years, what happened? What did we do? What did the governments do? What did Facebook do? What did we as a society do?

Claire Wardle: This is such a great question. We took it seriously and by taking it seriously we slowed down. We will add a ton of friction into the system that will stop our basis instincts just sharing without trying.

Tristan Harris: What does that friction look like for example?

Claire Wardle: The great work by Nathan Matias, who's now at Cornell who's shown that the more you ask people, "Are you sure this is true before you share it," the more you put delays in.

Tristan Harris: Limiting the number of people you can re-share something to.

Claire Wardle: Yeah, all this stuff that we know, giving more credibility. For example, if you have heuristics about where this information came, you're more likely to, "Oh well are you sure?" We have enough evidence now about things that we know, but I think it's about slowing down. If we had more data, we had not rolled out new features without proper testing. There was real work with academics before stuff was tested in the wild. I don't think it's a million miles away from where we could get to. The last three years I've probably gone to 150 convenings about misinformation globally.

Tristan Harris: I think you're the information disorder conference queen, I've seen you at every single one I've been to.

Claire Wardle: Right, and there has been some very good conversations at some of these types of things. In these last three years, whenever we've talked about this problem is complex, it's going to require a lot of people being involved in the solution it's going to take time, people are like, "Yeah, but we just need Facebook to tweak the algorithm." We've had three years of people expecting a simple quick response.

Claire Wardle: One thing I would say, in the last six months is there's this recognition of wowzers, like this is going to get a lot worse before it gets better. It's probably going to take 50 years not to solve but to get to a state that's more humane and is actually not causing harm. In order to do that we need deep education of everybody at every sector, and it's going to require real cultural shifts.

Tristan Harris: I want to walk outside and I want to say, I have faith that the people around me are being embedded in a healthy information environment. Slower, more thoughtful, more reflective, more careful, higher friction information environments. My friend Eric Weinstein has this great saying that, instead of critical thinking we're going to need critical feeling. That we have to examine our own emotional reactions to the information that is presented to us. Not just information but to experiences that are presented to us.

Tristan Harris: I'm actually kind of hopeful in this weird way with Coronavirus, where we've been in this low trust amusement driven world. Well you better bet people are going to suddenly be concerned with what's actually true when it comes to the health of your family.

Claire Wardle: Disinformation works if it taps into fears about your own safety and those of the people you love. On this that's exactly, you can dismiss all this political nonsense but this is about real harm. You touched a little bit on emotional skepticism and there are great media literacy programs. There's a lot of money that's gone into, "Oh my God, we've got to teach the 13 year olds." I think we know now that actually some great research from NYU, it's actually over 60 the other biggest problem.

Claire Wardle: I wish that we could do more to teach people how to talk to one another about this. We talk about this in training but for example, if I go home for thanksgiving and I'm like, "Hey uncle Bob, couldn't help but notice you posted something on Facebook. It's wrong and here's a Snopes article that proves that you're wrong."

Tristan Harris: Doesn't work so well.

Claire Wardle: No, and we know this psychological worldview theory which is like you double down on your worldview.

Tristan Harris: People feel great when people tell them that your identity is wrong.

Claire Wardle: Yeah, it turns out. We talk in training about how actually you're like, "Hey Bob, couldn't help but notice you posted that but I've been thinking a lot about this. Why are people trying to manipulate our communities? I'm watching this happen, people are trying to divide us. What do you think? Why are people doing this Bob?" I know that's a very simple explanation but the language of we and us is not the language that journalists and fact-checkers and researchers like to use. We have to be better at teaching one another how to slow it down, how to get people to take responsibility for the information they share.

Claire Wardle: I mean you and I have been talking for how long now? We haven't talked about people, the users and how we are being weaponized. We can add flags to Facebook and add more label, all the rest but actually if we don't stop my mom or whatever sharing it, then we're in trouble. We don't talk about that at all and so teaching people how to reverse image search a picture, yes fine. How to read a headline, fine, but we haven't talked about the psychology of us and we should be doing more of that.

Tristan Harris: There's this campaign that was brought to my attention by someone who actually helped create some of the major platforms. The theoretical name for the campaign was, We, The Media instead of we, the people. In this new world of user generated content, you and I are the journalists now. Even though we don't even think of ourselves that way because we are the information oncologist. Instead of someone who went to journalism school, knew the training, know that you have to ask for the opposite opinion before you publish the thing or at least make sure you reach people with corrections, all of those kinds of basic rules, we're not operating with any of those rules.

Tristan Harris: When you do the Indiana Jones swap of we have this media environment, where you had all these people who studied certain norms, standards, ethical codes producing information with certain flaws. Then you do the Indiana Jones swap into this new environment, where it's each of us are now the unpaid gig economy attention laborers who are driving around attention for free. Using our own vanity and ego and narcissism coning to get as much attention as possible. Each of us are essentially the information providers, but we don't have the responsibility or the norms that protect us from making mistakes.

Tristan Harris: If I do something that's misleading, that never shows up in my reputation. Imagine if like next to your... On Twitter, it shows you the number of followers someone has, the number of people that they follow. What if it had the responsibility score? People aren't going to like this, sounds like China, but if you imagine there is some notion we could agree on some set of values. That set of values would accumulate into a reputational score like a credit score. Not based on true or false, but just here are some standards of open standards for I don't know what those things would be. This would be kind of a, we'll work it out in real time. How would technology adapt to support something like that? You can imagine that being built into the design of products.

Claire Wardle: Yeah. I say this sometimes which is, 30 years ago we could be at a party and you could be very drunk. I could let you get into the car and go and I'd say to my friend, "I hope Tristan gets home all right." Now I would have to take the keys away from you because society has said, I cannot knowingly let you get into a car drunk because that's not appropriate for society. With this how do you say, "Wow, like Brian, last week you posted at least three false things on Twitter. It's kind of embarrassing mate."

Claire Wardle: We have to as society say, "We have to take responsibility for what we share." Again, I use this sometimes it might seem simple but like littering. Every time somebody shares something on Twitter that's false, it's like throwing a can of Coke out the window, what's the worse? Somebody's going to pick it up. We have to say, "Yeah, but if we all do that, we're in trouble," because I think the audience has been completely absent in these conversations. You go to any of these convenings like what can the government

do? What can the platforms do? What can educators do? Yeah, it's like we have to take some responsibility and right now I don't think there is any responsibility placed on us.

Tristan Harris: I think the reason people like myself we often turn to the technology platforms to say, "You've got to fix this," is because they operate at scale. The tech platforms are the vehicles by which you would distribute this education. If you were running Facebook right now, I know people ask you this all the time, or you're running Twitter or YouTube, what would be the way you would use that distribution vehicle to even enhance personal responsibility?

Claire Wardle: The worry about, "Hey Bob, you shared three false things last week," is it's really difficult. We don't have the AI systems right now to be able to automate that process. A lot of these stuff is the gray stuff, so it's actually harder to do. I mean a couple years ago the Guardian newspaper started adding yellow labels to say, "This is from 2014."

Tristan Harris: I loved that, yeah that's a great example.

Claire Wardle: It's like such a simple intervention and so of course there's a lot of hoo-ha about, would labels work? Maybe they're going to backfire, blah, blah, blah, but there is research now that actually the satire label would make a difference. There's 82 different satirical sites around the world, how many do you know? Probably The Onion.

Tristan Harris: Right.

Claire Wardle: The absence of those heuristics' means that Bob, you can't really blame Bob if nobody's helping him here. If we build in friction, we add context, then I think...

Tristan Harris: We do the Google meta tags things so that they can publish the article with the headline, that's about the truth instead of the false thing. The false thing is still getting picked up.

Claire Wardle: Yeah, all those kind of extra things and a lot of this is, how can we research in real time whether or not this is having unintended consequences? I mean the labels' thing, I'm really obsessed with at the moment, we're partnering with partnership on AI for a research fellow for six months to say, "Can we have a universal visual language around manipulated media?"

Claire Wardle: When the drunk Nancy Pelosi video appeared, some people called it manipulated, some people said doctored, some people said transformed. As an audience I don't really know what's happening here, so can we have a joint visual language that doesn't say, "That's a bad video," but we are in some way saying, "There's been an addition." Like how can we help the audience know what's happening? I'm interested in not media literacy campaigns or education, but how can that be baked into the platform in useful ways?

Tristan Harris: I love this, I think it was five years ago I had a side project that instead of fact-checking we called it frame check. I think we do need as an industry a common language in vocabulary on just what is the difference between the word distorted versus

manipulated? What's the difference between to steer, to guide, to persuade, to influence, to seduce? They're are on different dimensions of the degree of control that you have and the degree of asymmetry between one party and the other. How much does one know about the other party's manipulation?

Tristan Harris: We need this common language for the subtler terrain of how the mind if being influenced and persuaded. This would be a great thing to have baked into the common tech industry, because I do think it's almost like a missing component. We want to import this subtle human mind humane framework for how we work.

Claire Wardle: Yeah. I mean all jokes aside, I did my PhD back in 1999 in Communication and I was like, "Oh Mickey Mouse degree Claire." Now I'm like it turns out it was all this, it was all about framing, priming agendas, all that stuff. I'd say bad actors although I don't really like that phrase, but they have really good at psychology and emotion, really good.

Tristan Harris: To your point about disinformation that all these kinds of strategies have a kernel of truth, so the voters' suppression. I'm not saying, "Don't vote today," I'm saying, "Oh the lines are long today." Is it wrong to say, "The lines are long today,"? I mean what is long, what's the official definition of long? Is it two miles long? It's all arbitrary, but I'm creating a suggestion that maybe might be kind of hard. If you're feeling busy today, maybe it's not worth voting and that subtle ability to persuade.

Tristan Harris: I'd love to see a full-on cultural transformation, both in journalism and media and in technology and anyone working in the field of communications. Where we stop talking in the language of speech and we start talking in the language of subtle cognition.

Claire Wardle: Those of us who are pushing quality information we are dreadful at it.

Tristan Harris: Right.

Claire Wardle: We are rational, we're all about facts and it is an asymmetrical playing field. I actually did a talk last week at the National Academy of Sciences and it was the ugliest slide deck I've ever created. I basically just created meme after meme, after meme and I said, "This is how your adversaries talk to each other and this is how you talk. Here's your 187 page PDF with an image of a dripping needle on the front cover. That's not how this works." Of course there was laughter around, but there was a recognition afterwards which was like, we're really bad at communicating.

Claire Wardle: Just the other day, the WHO put out a leaflet about Coronavirus and they did exactly what you're not meant to do, which is like myth, myth, myth in big letters. Then underneath in small letters the truth. You should never include the myth.

Tristan Harris: What's the whole thing is that you're doing a lie sandwich, we're doing the lies more often than we're doing the truth.

Claire Wardle: How are we in 2020 in the middle of a Coronavirus crisis the WHO, who are an amazing organization, who hasn't taught them how to effectively push out this kind of

information using emotion but in a way that's compelling? Not dumbing down, but it's crazy to me.

Tristan Harris: Claire before we go, what can people do to help your work? I know that you are on the ground on the front lines of these things. I know that psychologically it's hard. I know that the lifestyle it's hard. I know that you probably lose a lot of sleep. What kind of support do you and the organizations that you're most working with need help with?

Claire Wardle: Particularly people who listen to this kind of podcast, there has been this creation of this divide between the tech press and the platforms. Which means that understandably there's this rejection of wanting to partner and work together. My sadness here is that, there are many of us who work on the ground around the world who really have got things to offer. Sometimes just a quick phone call or just coffee or breakfast, I just love some moments there, which is without signing an NDA, nobody's trying to get a got you moment of journalism that's one thing.

Claire Wardle: The other thing is to recognize that this work is, let's just talk about it for a second. I mean I spent the majority of my time fundraising as opposed to doing what I'd want to do. I say to some people, I feel like I've got two years to save the planet and that sounds crazy and insane and maybe over the top.

Tristan Harris: No, tell people why that's true. I think people really don't understand. Why is it true that we have two years to save this?

Claire Wardle: I started doing this work 10 years ago and I would stand in a room with BBC journalist and be like, "Don't worry too much about this, but just to let you know, during breaking news events, there's a couple of hoaxers who are probably going to try and manipulate you." That was 10 years ago. Now I stand in rooms with the same journalists and be like, "You might have gone on hostile environment training previously when you were about to report for the Middle East. I'm about to give you hostile environment training because the way that you work now on the internet it is a hostile environment."

Tristan Harris: It's a hostile epistemic information environment.

Claire Wardle: "Let me tell you how you protect yourself, how to stop yourself being doxed, how to stop yourself from being harassed, how you stop yourself from being manipulated." I see the speed at which this is happening, in two years' time this country will be fully polarized. We will have two different sets of media, nobody will believe anything from anybody and I do think that there is still hope. We cannot keep talking at convenings, we can't keep talking at podcasts about what we're going to do. We could have done this podcast three years ago and said almost exactly the same thing.

Tristan Harris: I know you and I could, I know.

Claire Wardle: We don't need a UN agency for disinformation because that's going to take too long to set up, but we need to work quickly. We need to be agile. Like my sister creating a Cross Check project around journalists collaborating, when that never would have happened

before. What do we do with this moment of inflection? I know everybody says this, but what do you want to look in the mirror and see? I just don't think we're taking this serious enough and I think Coronavirus might be the thing that all of a sudden makes people go, "This isn't a joke."

Tristan Harris: Yeah. Claire, thank you so much for coming on and I'm just a huge fan of your work, please keep doing what you're doing and we all support you even if you feel alone sometimes.

Claire Wardle: Thank you very much.

Aza Raskin: Your Undivided Attention is produced by the Center for Humane Technology. Our executive producer is Dan Kedmey and our associate producer is Natalie Jones. Noor Al-Samarrai helped with the fact-checking. Original music and sound designed by Ryan and Hays Holladay and a special thanks to the whole Center for Humane Technology team for making this podcast possible.

Tristan Harris: A very special thanks to the generous lead supporters of our work at the Center for Humane Technology, including the Omidyar Network, the Gerald Schwartz and Heather Reisman Foundation, the Patrick J. McGovern Foundation, Evolve Foundation, Craig Newmark Philanthropies and Knight Foundation among many others. Huge thanks from all of us.