Tristan Harris:

Structurally, we don't have a lot of information about how things operate inside the major tech platforms. Most people who lead these companies sign nondisclosure and non-disparagement agreements, so they can't speak publicly about their experiences.

Tristan Harris:

But Sophie Zhang was different. As a data scientist at Facebook, Sophie stumbled across a pattern that was going underneath the radar. Leaders of small countries like Honduras, Azerbaijan and Albania were exploiting a loophole to make them appear more popular than they actually were. In 2020, on her last day at the company, Sophie blew the whistle. She posted a 7,900 word memo to Facebook's internal site about how foreign leaders were exploiting the platform. She then turned down a $64,000 severance package in exchange for the privilege of speaking out.

Tristan Harris:

Today on Your Undivided Attention, Sophie shares with us how she came to feel responsible for the political futures of countries that she'd never even visited, and what she thinks can be done to prevent that kind of exploitation going forward. I'm Tristan Harris.

Aza Raskin:

And I'm Aza Raskin.

Tristan Harris:

And this is Your Undivided Attention.

Sophie:

So, I joined Facebook from the start to intending change it from the inside. I mean, I was upfront with them about that. I told them, "I don't think Facebook is making the world a better place." That's what I told my recruiter, "That's precisely why I want to join." If the company is perfect, there's nothing to fix. I'm joining it precisely because it wasn't perfect, so that there are things to improve, so that I can make a difference. And frankly, I hadn't anticipated how much of a difference I would actually be able to make.

Tristan Harris:

What was their response to that? When you said, "I don't believe Facebook's making a better place and I would like to make a difference!"

Sophie:

I think they were like, "You'd be surprised how many people at Facebook say that." Because ultimately Facebook is, or at least historically was, a very open company. Open dissent about the company was allowed and encouraged. We had weekly question and answers with the CEO. I don't
want to make it sound like the entire company was dissidents. If you think Facebook is amazing, you are more likely to want to work for Facebook.

Sophie:

But overall Facebook was a rather non-hierarchical company. It presents itself as very open. There was significant amounts of truth to that because dissent was allowed, and to some extent, accepted. They certainly wasn't [sic] shy about criticizing the company from inside while I was there. And as well, employees were able to easily speak with everyone else at the company, which I understand is much more difficult at many other companies.

Sophie:

With that said, I know Facebook has made changes since I left. My understanding is that they've restricted discussion of non-workplace related matters. Obviously I don't work at Facebook anymore, I can't say from firsthand experience, but I'm talking about my experience during my time there, rather than since I've left.

Sophie:

And it was also not very hierarchical. I was an IC4 at Facebook, which means nothing to the vast majority of people who are listening to this. So I'll translate that. I was literally one level above a new hire straight out of college.

Tristan Harris:

That's called an IC4, you said?

Sophie:

Yes. IC stands for individual contributor, as opposed to a manager or something. I was not anyone important within the company. There were probably a thousand people, data scientists, with my rank. And they would only want more because Facebook doesn't have a ton of data scientists. It's predominantly engineers, product people, et cetera.

Sophie:

But even though I was very low-level, I was able to interact with leadership, worked on extremely important matters. I personally briefed Guy Rosen, who's the company Vice President of Integrity, on what I found in Honduras. And I want to highlight how unusual that was. It would be like an Army Sergeant briefing Kamala Harris on something. It would never happen. And if it did, it would be emblematic of something very unusual going on. And so that is one strength of Facebook, though I'm sure they're second guessing it very much right now.

Sophie:
But, there was this culture of openness and non-hierarchical matters, that I was able to jump in and work on these problems and try to fix them, and have quite extraordinary access relative to my actual rank and title. We had lots of slogans, every company has its slogan. And at Facebook, one of them is that "nothing at Facebook is someone else’s problem."

Sophie:

And the idea is that, if you see something wrong, you should try to fix it. You shouldn’t assume that someone else is fixing it and was taking care of it. And I could always point to that and say, "See, I'm just trying to live by the motto and the logo." And just to clarify from the start, because I want to be clear about what I did work on and what I don't, to avoid confusion, that I worked on inauthentic activity. And by inauthentic activity, I mean, fake accounts that are conducting various activity, which is separate from misinformation and separate from newsfeed decision-making.

Sophie:

So I was on the fake engagement team. And so by fake, I mean, fake accounts, hacked accounts, self-compromised accounts, in which people went and they gave over their credentials to some nefarious actor; and by engagement, I mean, likes, comments, shares, et cetera.

Sophie:

There are teams at Facebook who work on civic integrity. There are teams at Facebook who work on what’s called coordinated inauthentic behavior, which is what... the investigations they do for things like Russia 2016. My team was not in either of those areas. It was largely a spam team. And by spam, I mean, just the more general definition of things that are low quality, repetitive, not individually harmful, but harmful in large quantities, in aggregate.

Sophie:

But you can also see how that was my ticket and excuse to working on matters that I considered important in my spare time. I was able to work on increasingly more sophisticated political activity and work closely with the teams that did work on coordinated inauthentic behavior. So I'm going to use an analogy here, and that's the difference between an FBI agent and a police man for a small village in say, Idaho. Sorry, Idaho.

Sophie:

The average person might think they're vaguely similar, they're both cops, but these are jobs that are very different in scope in organization, in importance and training. And if we compare the two of them out loud, the FBI agent might get a bit offended. And so my situation was essentially that I was some local village policewoman, not very important, except I kept finding things that went up to the FBI, in this analogy, which would be very unusual and would have political consequences, in internal politics. Because on one hand, the FBI is happy that they're getting all this work
done for free. They don't even have to assign anyone to do these investigations. On the other hand, they feel a bit embarrassed or upset, they're being strung up. And they may be unhappy that she's giving them extra work, "Who is she to come over and give me more work? I mean, she's not my boss. She's not even in my organization."

Sophie:

And meanwhile, her own village police chief is maybe a bit annoyed that she's doing all this FBI work instead of working on village crime, which is where she belongs. And so that's the analogy I would use to try and describe the situation. Because fake engagement sounds very serious, but most of it is not that severe and it's focused on high-volume scripted activity. And I was essentially petty crime department. That was my team. That was my role and my job.

Sophie:

But my job was defined widely enough that I could argue that it made sense for me to look into more important, or volume, cases in my spare time. I never got training. To the extent, the oversight and guidance that I got was, "Stop doing this and work on your actual job." Everyone knew what I was doing. I talked about this up to the company Vice President, it wasn't a secret, but it just wasn't officialized or regularized. I was essentially running a shadow integrity program.

Aza Raskin:

The way I'm understanding it is that the bulk of the problem, when you just think about fake engagement, is this sort of low quality, unsophisticated spam; there's like lots of these little petty crimes. Then there's the FBI that's really focused on working on high sophistication, high impact, fake engagement, things that affect, you know, world elections.

Aza Raskin:

But in between here, there is unsophisticated, there's high sophisticated, but then there's semi-sophisticated that sort of sits between the two, and no one was looking at that. And that's what created the space for you as a sort of town cop to move up and be like, "Hey, an entire swath of activity is being missed."

Sophie:

I think that was part of that. And I think the FBI agent would be annoyed at you, in that analogy, just like the teams would be annoyed at you in this analogy because you're like, "That's not fake engagement. Fake engagement is low level." Just like if you compare an armed bank robbery to shoplifting, they're like, "This isn't shoplifting, this is much worse." And so it's definitely the case that teams that were focusing on this were responding to cases that were more important.
Sophie:

But I would also highlight a different aspect, which is that they're also responding based on PR considerations, and what I mean is that they're primarily reactive rather than proactive, at least in my time at Facebook. They weren't usually going out and looking for things, and if they did, it was mostly in countries like the United States. Rather, a lot of their time was spent looking at outside reports from news agencies, from opposition groups, from NGO's saying, "This is a weird going-on. Let's take a look at it."

Sophie:

I would say the area that I really pioneered is that I was going out and looking for it on my own. There wasn't anyone who pointed me to this. I don't think if you asked anyone, "Pick two countries in the world that you think are most important." I don't think anyone would respond, "Honduras and Azerbaijan," quite frankly. They're an awkward combination.

Sophie:

But I think in the case where, as you're looking into Honduras, they weren't even particularly trying to hide.

Sophie:

I literally just pulled up the person's Facebook page, I didn't know who Juan Orlando Hernández was until I Googled him. "Oh, wait, he's the president of Honduras. Great, I guess." And then I opened up his Facebook page. I wanted to take a screenshot for a PowerPoint, and I just wanted to open up the people who had liked him and take a screenshot. And then I stopped because I realized something very surprising to myself. And that was that most of the people who were liking these posts, they were not people at all, they were pages pretending to be people.

Sophie:

And so pages are a Facebook feature that are meant for businesses for instance, for celebrities, influencers, politicians, et cetera, people who have a public profile. They are a public-facing thing that's controlled by individual users, but this was a loophole at Facebook. There are smart people who are at Facebook, who go out and look for fake accounts. There was no one at Facebook, and to my knowledge, still is no one, who goes out and look for pages pretending to be people.

Sophie:

An analogy I'll use is that for instance, jaywalking is illegal in much of the United States, but it doesn't mean that the cops go out and find people jaywalking and stop them. And so this is a loophole that was used by adversaries throughout the world. It was used by the governments of Honduras and Azerbaijan.
Center for Humane Technology | Your Undivided Attention Podcast
Episode 37: A Facebook Whistleblower

Tristan Harris:
So some listeners might be thinking to themselves, "Well, how consequential is it?" So if I can run 200 accounts and impersonate some public perception and sort of steer conversation on a few posts, this sounds like a pretty negligible thing. But yet you in your memo said that you felt like the stakes of this work were so high, that you were losing sleep. You used the phrase that, you felt like you had blood on your hands. Not in the sense that you caused something, but in the sense that, whether or not you would intervene would be consequential for an entire country or for an entire opposition group or for entire minority group.

Tristan Harris:
How would you talk about how significant this set of issues really is for those who are skeptical that this has any influence at all?

Sophie:
So I want to be clear right now. That it's always hard to figure out the uncertain consequences of these sorts of indirect effects. You can't say, that this newspaper article caused XYZ who happened by changing X opinions. I'm not an expert on politics. I'm not an expert on public discourse. I'm not an expert on public relations. But if anyone is an expert on politics, public relations, and discourse in their nations, it's the presidents of those nations. Because you don't become the leader of a nation by being stupid and throwing money down a hole.

Sophie:
If you stay a national president for more than a decade, you have to be very good at maintaining the support of the people and the support of the important people. And so I'm not the expert. But it's very clear that those people who are the experts have decided that this is worth the money and expenditure. I caught the governments of Honduras, I caught the governments of Azerbaijan red handed. They weren't even trying to hide.

Sophie:
And to me, that's very telling because, I mean, Honduras, Azerbaijan; these are not the wealthiest countries. Their presidents do not have a ton of money to spare. These people genuinely believe that this was worth the expenditure, the effort, paying the people to do this full-time, and the risk of being caught. And to me, that's telling. That's suggestive. The people who are the experts have spoken.

Sophie:
The analogy I would use was the Russian interference on social media in 2016, which I think was very shocking and jarring to a lot of people. And it's hard to say how much of an impact that actually had. I don't think anyone can say for certain. And I think that there are always arguments that it wasn't that decisive, but I think it was still very shocking to the psyche of the nation. And what I found in Azerbaijan and what I found in
Honduras was, I think, rather worse. Because in Russia 2016, that was a foreign nation. You expected that from Russia.

Sophie:

For Honduras, for Azerbaijan, it was their own country. It was their own president, their own government that were doing this to them. And that, to me, that made it an order of magnitude worse. When people imagine the sorts of consequences for this sort of behavior, I think they usually jump to the conclusion that this is intended to persuade, or et cetera. But I would rather focus on a different aspect. And that's the perception of popularity, the perception of public opinion.

Sophie:

When it became clear how much I was able to do in my spare time, I think I took a step back and decided what my priorities were and what my standards, what my moral compass was. For instance, I decided from the start that I would never act unilaterally. I would try very hard to avoid being judge, jury and executioner. I would only act in the role of investigator. I would always let other people confirm. I would always let other people sign off to take action; I would always let other people do the actual enforcement, as long as I could help it.

Sophie:

Because I mean, honestly, I could have gotten away with asking forgiveness instead of permission in a number of these cases probably. And Facebook will say, and they're technically correctly in doing so, that I was never the only one acting. Like I said, there were always other people signing off. But in practice, I never received answers of, "No." It was always, "Yes, let's do this." Or silence, no answer. And whether I got a response, as far as I could tell, was entirely dependent on how much I yelled and made noise about it and et cetera, that got a decision to be made, which was invariably, "Yes, let's take it down."

Sophie:

In some cases, it happened overnight. In other's it took more than a year. In some countries it hasn't even happened at all. Most people try to distance themselves from the consequences of their work. They try to keep it at arm's lengths. They do what they need to, to sleep at the end of the night. But in my case, I thought that there wasn't anyone who would make those decisions, would take those responsibilities, if I didn't. And so that's why I decided to take that onto myself, I suppose.

Sophie:

Ultimately, it's hard to say how much of a difference I made. It's hard to say how much a difference any of this made. That's the crux of the issue. I'm trying very hard to be cautious and be realistic of what we know and what we don't know. And for indirect impacts, it's always impossible to say what the actual consequence is.
Tristan Harris:

One of the other topics that comes up in our discussions around this area is what a friend of ours has called the gradient of privilege. That, as you mentioned, the number of days that it took to action content on these other countries that are sort of further down the food chain in terms of priority. You mentioned in your memo that —

Sophie:

So just a tiny clarification. I would clarify that this is actioning users and behavior, not content.

Tristan Harris:

Yeah. Excuse me. Users and behavior. And so in the Guardian article that highlighted here, this is, how many days did it take Facebook to address inauthentic behavior after an employee flagged the case. In Poland, it was 1 day; in the Philippines, it was 7 days. In Taiwan, it was 11 days. India, 17 days.

Sophie:

It was only seven days after I found that the US connection there's a separate part for the overall network. But anyways. Yeah, it depended quite a bit on circumstance. It depended quite a bit on who the person responding was, because there's a lot of variants. What happened in Poland, wasn't that Poland was number one based on Facebook's list, it was very silly. The employee looking into it was a Polish person. He was like, "Oh my gosh, this is happening in my country." I mean, I'd flagged that the day after Christmas, I woke up the next morning to see that he had already taken it down. Actually the policy people were a bit upset.

Tristan Harris:

You are bringing up such an important and fascinating point here. Because part of this issue is about representation. If Facebook had employees that were... And there's still a question of, "Why should they be the ones playing God for their own country? And does one person who's Polish help... best represent what should happen in terms of what accounts you take down or not in Poland?" But certainly the fact that he or she cared about that, is interesting.

Sophie:

It's certainly the case that, I think, most people care more about their own nation. That's true for every country and for region, like the United States cares more about Britain and Western Europe because of cultural affinities, because of language affinities for the United Kingdom. And so Facebook's employee base is, mostly, comprised of a mix of American/Western European people, plus Indian and Chinese employees here on [H1-B] visas. And so Facebook does face employee pressure on these points, because I mean, people care naturally about their own countries.
Sophie: And I think I’m very unusual in that I came forward, I quit based on countries that I have no connection to. And so to a large extent, Facebook has faced a lot of pressure from its employee base on countries like India and the United States, in some cases more than the PR pressure, because Facebook employees are very difficult to replace, especially if they all quit en masse.

Sophie: And that’s essentially a dynamic that privileges certain countries over others. But even when countries are prioritized, that priority and treatment is often not what people within those countries would prefer, I think. Because like India is a very large country and it is a priority country in Facebook’s metric. It’s as important as the United States in Facebook’s metrics, they are on the top tier.

Sophie: And so in India, when I found inauthentic activity that was being ran for political activity, Facebook was much quicker to take it down, with the exception of a single case, where we had gotten permission to take it down, the sign-off to take it down, we were about to act and suddenly we realized that the network was run by an important Indian politician. This was a member of the Indian parliament, the Lok Sabha, who was honestly so arrogant was silly was stupid, that he hadn’t even bothered to use a VPN to run his fake accounts.

Sophie: And as soon as that was realized, suddenly everything stopped. I never got an answer or even acknowledgement. I never got a yes or a no. I kept asking. And you can argue that, “Okay, you send someone a message who don’t respond. Maybe they just didn't see it,” because I mean, sometimes people don’t respond to emails, they miss things. But when you keep asking them, they keep not responding. When you’re already in conversations with them about other matters, and you talk about those matters perfectly fine, then you add in, “Well, since you’re here, what about this?” and they ignore you, and this keeps happening, it’s very clear that something unusual, is going on at that point.

Sophie: Ultimately, I guess Facebook decided that keeping this person happy was more important than protecting the sanctity of the world’s largest democracy. And I disagreed with them strenuously in that regard. And I think that most Indians would also have a different opinion because I mean, importance, again, and decision-making, is taken through the lens of Facebook’s self-interest, because of course, Facebook is a company.
The US probably gets... And the US elections probably gets the most amount of Facebook’s resources, attention, bot scanning —

Sophie:

They absolutely do, but a separate point I want to raise is that Facebook has more leeway to act independently in countries that are less important.

Tristan Harris:

Very interesting.

Sophie:

And what I mean is that, Facebook was fine to fine the governments of Honduras and Azerbaijan at my say so. If I had caught the government of India doing this, they would have absolutely never have done the same. They would have had long discussions and then found out some reason to say why they shouldn't act or something like that.

Sophie:

The government of India is quite powerful. They can threaten to arrest Facebook employees, they have threatened to arrest Facebook employees, if Facebook doesn't go along with their desires. If the government of Honduras bans Facebook, then that hurts Honduras far more than Facebook, quite frankly. Its own citizens would be very unhappy, but India does have a lot of leverage and power over companies like Facebook. They can demand that Facebook essentially send employees to be based in their own nation, to be served as potential hostage trades, essentially.

Tristan Harris:

That's actually what I understand they're proposing right now, that India has actually asked to make ... Not asked, they're passing, I think, laws to make sure that social media companies have physical, on-the-ground representation from employees. And then to, after that, to impose criminal liability if the companies are not acting them. So you're actually talking about a very real scenario of essentially taking hostages so that you can get your way.

Sophie:

Yes, absolutely. Because on one hand countries that were less important, these got less priority and attention, but the lack of attention also meant that decisions could be made with less political interference. That things got made even when the governments disagreed. Sometimes lack of attention can be a benefit. It's a double-edged sword.

Tristan Harris:

In a sense, there really is a playing God around the world and sometimes that God is beholden to a certain power or interest, and then sometimes
that God can just be... So if he’s on, making a decision on behalf of some of these other countries, and I can imagine just the feeling of weight and responsibility that is reflected in your letter.

Sophie:

Certainly. The usual defense that Facebook has is that it's hard to find things, but I think part of why precisely my story and my work was so powerful and spoke to so many employees, was that I had done all the hard work for them already. I had already found the bad people, and yet they sat on it.

Sophie:

They aren’t trying to dispute anything that I have to say. They can’t because they know I’m speaking the truth and because I’ve documented all this extensively. In Honduras, what we found from the study, it was very obvious what it was going on and especially after I realized that one of the people controlling hundreds of these fake pages pretending to be real people, was one of the page administrators for the Honduran president. This was someone who had been specially trusted and charged with making statements on the Honduran president’s behalf, to have access, to say things for him. And that’s one of them.

Sophie:

And so it was very clear from the start that this was a troll farm that was run by the government of Honduras or someone authorized by them. I took this to the Facebook leadership in my presentation, like I said, and the response I got was very surprising to myself. And that was, a policy leader calling in from somewhere in Latin America, he told us that he wasn’t surprised. Because he had already been told socially, by people who ran social media for the Honduran government, that they had been tasked with running troll farms on the Honduran government’s behalf.

Sophie:

And at the time I was very shocked that they would just go out and admit something like this. In retrospect, I think that should have been my first red flag that he had been told this. I mean, they’d known about it already. I was very naive from the start. I thought, “Okay, I found this, I’d hand it over. They’d take care of it. Everything will be great. I’ll go back to my actual job.” Instead, it was the start of a Sisyphean two-year ordeal.

Sophie:

It took almost a year to take down the Honduran government’s operation, they came back almost right afterwards. In Azerbaijan, the Azeri government was also running their own domestic troll farm. It took more than a year for that to be taken down. And as far as I know, they’re both still going on. In terms of potential solutions that I would offer, what the companies can realistically do and might be vaguely in their self-interest, I’d suggest two separate things. The first of them is this.
Sophie:

So at Facebook and other at social media companies, I presume, but I can’t speak about them, just about Facebook from my own expertise. At Facebook, the people who are charged with adjudicating cases, of saying, "Do we act in this case? Do we take down this person? Do we take down that person?" is the same people who are charged with maintaining good relationships with governments and important figures, which creates extremely obvious conflicts of interest. And Facebook would argue this is, I mean, it’s a company, this is in its own self-interest, it’s important to, et cetera.

Sophie:

But I mean, it’s also the case that most news organizations are for-profit publishing and organizations, and most news organizations do have separation between the editors who decide what pieces get published and the people who are charged with having good relationships with governments. If a piece would make a government angry, then they’ll have the PR people handle it, but they don’t have the PR people say, "We’re vetoing it because it will make the government angry."

Sophie:

And so, the news media has managed to at least have that theoretical separation. And I think it would be beneficial for Facebook, the other social media companies, to have that as well, even if I don’t see it happening anytime soon. The second is more broadly just that I think many companies like Facebook that rely a lot on metrics, can be a bit short-sighted in their scope. And what they mean is that Facebook cares a lot about its numbers, about improving its profit, or et cetera.

Sophie:

But the issue with numbers is that if you only pay attention to metrics, if you emphasize them, you naturally de-emphasize things that can’t be measured. And I think Facebook has suffered in this area to some extent, because it’s been very bad at proactively managing its PR. I don’t think it’s controversial to say that, the existing Facebook public relations, the public sentiment on Facebook is absolutely terrible. And so Facebook does have incentives to get ahead of negative PR, to prevent problems before they become fires.

Sophie:

But I mean, preventative measures are much harder to justify than reactive, because it’s much easier for a fire department to say, "We put out 10 fires last year." They can’t go and measure, "We had fire safety codes that prevented a hundred fires from happening last year," because I mean, how do you even measure that? You can’t. And so the first seems tangible. The second seems wishy-washy. And so I think Facebook has difficulties with that to a large extent, because frankly, Facebook is a company that acts in its own self-interest.
Aza Raskin: So you left in 2020, yes?

Sophie: I was fired in September of 2020.

Aza Raskin: So you've been through a lot. You've seen a lot. You've blown the whistle. I'm curious for the other Sophies that are still within Facebook or within the Twitters or the TikTok's, what advice would you give them?

Sophie: I don't want to tell people what to do. Different people have different considerations. I don't have to feed a family. I don't have children to support, or starving relatives that I need to support back home or something. And so I don't want to begrudge others the decisions they have to make to keep the lights on for themselves. Everyone has to sleep at the end of the night and how we achieve that is up to themselves.

Sophie: But for the people who come in idealistic and do want to make a difference, I would tell them to think very carefully about what your goals are, what your motives are and where you would draw the line, because I think, working within the system to change things is difficult. It's very powerful because you have an extraordinary amount of access, but at the same time, people naturally like the people they work with better, they get used to the situations they work with.

Sophie: And it's easy for people who come in from the outside who wanted to fix the system, to get used to the way things work until their supporters feel betrayed. And so I would ask people to be very clear about what they want to achieve, what their goals are, how things would be received upon and seen by the outside world, and what your own priorities are, what your goals are. I would tell them that as a new employee, I managed to catch two national presidents red-handed and made international news on multiple occasions. That's what I did. What can you do?

Tristan Harris: Do you want to stand in solidarity with tech workers like Sophie? Join the Center for Humane Technology and a coalition of other aligned organizations, for a special event to build solidarity among tech workers and tech users on Tuesday, July 27th. For more details, visit humanetech.com.

Tristan Harris:
Your Undivided Attention is produced by the Center for Humane Technology. Our executive producer is Stephanie Lepp. Our senior producer is Natalie Jones. And our associate producer is Noor Al-Samarrai. Dan Kedmey is our editor at large. Original music and sound design by Ryan and Hayes Holiday. And a special thanks to the whole Center for Humane Technology team, for making this podcast possible. A very special thanks goes to our generous lead supporters, including the Omidyar Network, Craig Newmark Philanthropies, the Evolve Foundation and the Patrick J. McGovern Foundation, among many others.