SELECT COMMITTEE TO INVESTIGATE THE
JANUARY 6TH ATTACK ON THE U.S. CAPITOL,
U.S. HOUSE OF REPRESENTATIVES,
WASHINGTON, D.C.

INTERVIEW OF: FRANCES HAUGEN

Monday, November 22, 2021

Washington, D.C.

The interview in the above matter was held via Webex, commencing at 11:08 a.m.

AUTHENTICATED
U.S. GOVERNMENT
INFORMATION
GPO

Appearances:

For the SELECT COMMITTEE TO INVESTIGATE
THE JANUARY 6TH ATTACK ON THE U.S. CAPITOL:

[redacted] SENIOR COUNSEL
[redacted] CHIEF INVESTIGATIVE COUNSEL
[redacted] PROFESSIONAL STAFF MEMBER
[redacted] CHIEF ADMINISTRATIVE OFFICER
[redacted] COUNSEL
[redacted] CHIEF CLERK

For FRANCES HAUGEN:

ANDREW PETER BAKAJ, COUNSEL
LIBBY LIU, CEO, WHISTLEBLOWER AID

[redacted] All right. This is a transcribed interview of Frances Haugen conducted by the House Select Committee to investigate the January 6th attack on the United States Capitol pursuant to House Resolution 503.

Ms. Haugen, could you please state your full name and spell your last name for the record.

Ms. Haugen. My name is Frances Bordwell Haugen. Last name is H-a-u-g-e-n.

[redacted] This will be a staff-led interview. Members of the subcommittee may join, and, of course, may choose also to ask questions. My name is [redacted] and I am an investigative counsel to the Select Committee. With me today are [redacted] senior counsel and senior advisor; [redacted], a professional staff member; and [redacted], chief clerk, [redacted] chief administrative officer; [redacted] may also join. He is our chief investigative counsel. And if we can spell any of those last names for the court reporter's need, I will do them for you. I have circulated their names before.

Before I begin, I would like to describe a few ground rules where I will be asking questions for the staff, but [redacted] or others may also jump in. You and your attorney will have an opportunity to review this transcript and make sure that it is accurate. You are permitted to have an attorney present. If you do at this time, could counsel please state their name for the record, and maybe if you could also state your name and affiliation.

Mr. Bakaj. Absolutely. My name is Andrew Peter Bakaj. Last name, Bravo, Alpha, Kilo, Alpha, Juliet, Bakaj, and I am counsel for Ms. Haugen.

Ms. Liu. I am Libby Liu, L-i-u, and I am the CEO of Whistleblower Aid and representing Ms. Haugen.

[redacted] Thank you. There is an official reporter transcribing the record of this interview. Please, Ms. Haugen, please wait until each question is completed before you begin your response. We will try to wait until your response is complete before we ask our next question. Our reporter cannot report nonverbal responses, such as shaking your head, so it's important that you answer each question with audible, verbal response, and speak slowly, as I was just not doing.

If you need me to repeat a question, please say so. If you need to talk to your lawyers, please say so. If you need a quick break, give us the word, that's why we'll take a quick break. If the reporters have clarifying questions, they may interrupt us to make sure that they can accurately capture our statements. We ask that you provide complete answers based on your best recollection. If the question I am asking is not clear, please ask for a clarification.

If you do not know the answer, please simply say so. I also want to let you know that it is unlawful to deliberately provide false information to Congress, even though today we are not proceeding under oath.

The plan for today is we will try to get a good chunk out of the way before taking a break probably around 2:00 p.m. Eastern for 20 to 30 minutes. We will then resume and get through the remaining material. I plan to start with some basic background information about you, your scope, your work there before proceeding to ask some specific questions about some of the documents you provided.

Lastly, I hope to wrap up with some questions about Facebook's structure, any clarifications, and follow-up related to any new insights that your testimony has provided. Are you clear on that?

Ms. Haugen. Yes.

[redacted] And do you have any questions?

Ms. Haugen. I don't. Actually, I just have one question. I need to grab one last thing. I just realized I forgot something. Can we have a 30-second break; I am going to run and grab it.

[redacted] Okay. We are not going to go off the record.

Ms. Haugen. [redacted] So thank you for your patience.

[redacted] If it becomes disruptive, [redacted].

Ms. Haugen. Yes. [redacted] Thank you. Cool.

EXAMINATION

BY [redacted]

Q Let's go. So let's start with some work and education history. What is your post high school education?

A I attended Olin College at the end November actually. O-l-i-n. It is in Needham, Massachusetts. It's a brand-new college that focuses on interdisciplinary engineering education. I have an MBA from Harvard. I attended undergraduate from 2002 to 2006 and graduate school from 2009 to 2011.

Q What did you study at Olin?

A I have an electrical and computer engineering degree, a B.S., and my MBA is in general administration.

Q And after school, where did you work, your first job?

A My first job is at Google. I worked on ads. So I did a number of jobs during the course of my time at Google, so I was in a rotational program. Once I work on ads, I worked on ads reporting, which exposed me to lots of different ad types within Google. I worked on Google Books, including the lawsuit. They are the largest class

action lawsuit in U.S. history. Google paid for my MBA, and I came back worked on Google+, which was Google's attempt to build its own social network. And I founded the search team under Google+. And then I worked on a knowledge graph at the end. I was there from 2006 to 2014.

[redacted] so I wasn't working for a bit. Then I worked at Yelp, working on kind of product-oriented data mining. So we did things like, if you ever sort your photos for the money photo, that was like one of our first features. I ran ranking for the home feed on Pinterest. I was product manager there. I also ran ranking for related pins on Pinterest, which has like a -- it has -- a shockingly high fraction of all the content viewed on Pinterest is related to pins. I worked at a place called Gigster, which did enterprise software. And then I worked at Facebook from 2019 through 2021. So June of 2019 through May of 2021. And I worked on initially misinformation, specifically, civic misinformation.

So the third-party fact-checking program was in the main misinformation team. And our team worked on misinformation in times of crisis because it moved too quickly for third-party fact-checkers to be involved. And we worked on misinformation in places that did not have third-party fact-checking in general, which is places like everywhere like the United States and a few developed countries.

I worked on -- after that I worked on civic influence modeling, which is can you find influential -- civic influential people on Facebook, via how people interact with them or other signals. And we developed a system that was, you know, on load of like 80 percent effective, which was not high enough, unfortunately. And then, where did we go after that? Oh, and then I worked on the counterespionage team from November -- not November, October of 2020 through May of 2021.

Q Okay. Taking you back to your first job at Facebook, you started in June

of 2019, and you were with the misinformation civics team until when?

A Until February of 2020. And then I did civic influence from February 2020 through October 2020, and then countering misinformation from October 2020 through May 2021.

Q Got it. So starting with the misinformation --

A Oh, I have a few details on the countering misinformation. I also worked on a project that was pan the threat investigation org. So like we were supposed to build out tooling that would help with proactive detection, and recidivism detection across the intelligence org. So I got exposed to the information operations team, I got exposed to the dangerous organizations team, which is like terrorism and cartels. My last was, like, human trafficking and child exploitation. But I did not work on those problems, I worked on generalized detection software across those problems.

Q And so, after misinformation, your first role, what was your specific job title?

A Product manager.

Q Product manager. And what were the responsibilities of a product manager misinformation?

A So a product manager's in a general sense, their job is to talk to users and understand what other problems the users face, and then develop -- take those problems and kind of abstract them into other commonalities across them. Like, could you build something that would address maybe one on many of them. And then once you have a solution specified and a consensus built, you then break up that project into smaller projects and, like, help phrase them in like how do you then turn those projects into the most minimal version of the project. And then what will be the V2 of the project, maybe three of it.

And so, you're responsible for helping engineers to clarify their thoughts. So you

might have, like, when I worked at Google, one thing I frequently did was, you know, when I was an engineer I would sit in and -- tell me all the things they are interested in in this space. We put them on the board, and we would distribute them down into (A) there's clusters here. You know, you work on something here on work on something there.

Q Okay. And did you have any direct reports?

A The project managers, it's a weird relationship because you do not write, say, you know, the development plan for an engineer. But you do guide what the work of the engineer is. And so, I did not have direct reports, but in all of those cases, I had teams where, you know, I was the voice that guided what we did on that project.

Q So were you able to give instructions to engineers?

A Uh-huh. Not their managers. Like, usually, you have, like, a dual structure where somebody like guides what it is that you do, and then someone who is like responsible for like -- like, if one of the engineers is not doing their job, I would not be the one to fire them. But I would be the one who would say like, Hey, you know, we're doing this at this point, we're doing this at this point, we are doing this at this point. Does that make sense?

Q Sure. So where was misinformation located in the company?

A So there is two misinformation teams. There is the main misinformation team, which has 40 people on it, and that was third-party fact-checking. And that is under community integrity, or was back then. And then my team was under the civic ward, which was civic integrity. And I don't remember if civic integrity reported to community integrity, or if they reported directly to Guy Rosen, like, that org. But civic integrity sat in its own then, and civic misinformation sat within that org.

Q And Guy Rosen reports to whom?

A I believe Guy Rosen reports to Mark Zuckerberg. I am not entirely sure. I am pretty sure, though.

Q And Mark Zuckerberg is?

A The CEO of Facebook.

Q And so, in your second role, you said you were doing, I believe civic work?

A Yeah.

Q Were you in the same location in the company?

A Yes. So I still reported to the same manager, but I had a different -- I did an illustration of like this concept of like, there is product managers, and then there is engineering work. So I reported to Samidh Chakrabarti, who is the head of civic integrity. In my first team, those engineers directly reported up through the engineering org.

Q Could you please just spell Samidh's name?

A Oh, that one is hard. It's like S-a-m-i-d-h.

Q Okay.

A So that team sat in the -- part of my team sat in Menlo Park. Part of my team sat in New York. Then my next team was still in the civic integrity, or half of it was. So I had an engineer that was out of New York, who's still reporting to the civic engineering org, but I also had a researcher that sat out of Facebook connectivity. And, so, I was like the team. Facebook teams are extremely minimally staffed. So you frequently have teams that are between like one and four engineers.

Q And then your third job was the counterespionage team?

A So that team was --

Q Was that the same location within the company?

A So I was still reporting to my manager in civic integrity, but now my -- that

team was staffed by threat investigators, like threat researchers who worked within I3. I think I3 is integrity, intelligence -- I don't know what the third I is, but it's called I3, and it's where the threat investigators sit. And so my team there had a team of threat investigators and their manager, and they all reported up to -- what was his name? I would look up that name. I don't remember it exactly, but they had a separate org. And I think that org reported up eventually to Guy Rosen as well, but it might not have. I'm not sure.

Q Okay. And so, what made you want to work at Facebook in 2019?

A So I -- I think the people at Facebook are wonderful people. I have known a bunch of them. And I had a friend who had been radicalized on the internet in 2016. I don't know for sure on Facebook. I think it was more 4chan and Reddit, but you never know these things. And it had been an incredibly painful experience for me, because, like, this person I felt was hugely responsible for how [redacted]

Like, they would [redacted] Yeah, and so it was incredibly painful loss for me. And I didn't want anyone to feel that pain. And so I drained civic integrity, I drained civic misinfo. And almost immediately, it was very apparent to me that things were substantially worse than I had thought it was before I joined. That's -- yeah. That's why I joined.

Q And what is civic integrity's role at Facebook?

A Civic integrity's mission, I think, I had somewhere with them that they the document cash, there is a slide that would literally say civic integrity mission. The mission of it is to make sure that Facebook is like a positive influence in the world. So it's both in the scope of elections, especially in the scope of just general like, you know,

things like Myanmar, or like Ethiopia, like, preventing violence from escalating.

Q It sounds like you had a number of very interesting and different roles of civic integrity. So who within the company was civic integrity to work with?

A So civic integrity works with -- it's an interesting question of what does it mean to work with someone? So civic integrity has its own engineers, has its own data scientists, has its own researchers. But at the same time if something we did would impact, say, the news feed, we would have to go talk to people that were joining this feed and make sure they were okay with the change that we had done. And the -- I am trying to think what else is interesting. Civic integrity -- they work with country operations, which is describing documents as CO. CO is where all the moderators sit, and people who do operational roles. We work with the policy team. So I worked with the policy team a good deal in a counterespionage role. We also worked with -- who else did we work with? Give me one second. I'm thinking. Yeah, I think that's like the main stakeholders that we worked with.

Q And do you work with strategic response?

A I did not work with strategic response much. I knew they existed, because in my first role on civic integrity, civic misinformation, one of the things that we explored doing was Facebook as a policy of, like, not changing how the general system works until basically there is a crisis. So it means there is like a basically binary operation where it's, like, you're either in okay zone or either in, like, inflamed zone. And one of the things that we saw or were hoping to do was if we could develop a quantitative model for describing, like, how bad are things, we might be able to begin turning on some of these safety systems incrementally instead of asking like is the house on fire? Does that make sense? And so, during that period of time, I did work with strategic response in terms of learning how, like, that team operated at a high level, so that we were aligned in terms

of -- so we were aligned in terms of not duplicating work.

Q Just a quick look on this. Do you have any direct reports for that when you were the product manager?

A Yes.

Q And were you also a project manager?

A So product managers sometimes do some project management. So to differentiate between project management and product management, which are things people commonly get confused. Project managers take a goal that has been given to them, and maybe even like a list of things that need to be done, and then work with people to make sure that it's like a phasing, and that all people who need to be updated are, like, updated sequentially at, like, the right moments.

Product managers help to find like what is the problem to be solved, and, like, what is the solution that we're willing to pursue. And sometimes, that involves also doing some project management to support that role.

Q And how big was the integrity team when you were there?

A Oh, well, I don't know for sure. I feel like it would have been -- so, usually, actually, this is an interesting question. So Facebook has said before they have 6,000 people working on integrity. And I don't know what that possibly means. Like, it might mean that they're -- like when you ask for data on that from them, you need to have them make it by role. Because I don't think there is any chance there are 6,000 engineers working on integrity. That doesn't even seem plausible. And, so, I'm guessing, like, counting with the net, like, people who label data and people who -- I don't know, mainly low-level screeners on some system or something.

I think there must have been like under 1,000 people at counter integrity. I don't know that for sure. I know civic integrity had maybe 200 or 300 people.

Q And you mentioned engineers as being people who are particularly important?

A Yes.

Q There were 6,000 people that would have been some of the ones that was important to pay attention to how many there were?

A Yes.

Q Why?

A So the people who actually kind of -- so the only people who changed the products are engineers, right. So if you don't have the ability to change the product, you don't have -- like you are working with other levers that already exist. So if you say don't have a problem that is currently being solved, you end up a certain distance on your own. Like I said, you were trained detecting a new kind of problematic content, you could go a little bit about content. It's like, say, like, this is a good piece, or this is bad piece, that kind of thing. You could go about a policy about it. You could potentially -- there is certain kinds of things that are a little bit more off the shelf that you could do.

But in general, you can't solve that problem unless you have engineers. And so engineers are kind of the measure of how much effort is being exerted to solve the problem.

Q And all engineers work on civic integrity?

A I think on the order of -- I don't know it was more than 200, maybe 300. But we could check on that.

Q And how was civic integrity organized?

A Civic integrity is organized into pods when I was there. So at the very beginning, I think there was four pods. So it was, like, participation,

knowledge -- participation, knowledge -- I don't know what the other two were. So they worked on things like photo disenfranchisement; they worked on what they called social cohesion. So social cohesion is a euphemism. So when ethnic violence happens, or genocide happens, social cohesion is broken down. So they worked on -- I worked in the crisis response software. So if someone called rapid response, initially, and they run -- they used to call them war rooms, but they changed the name to be IPOCs. So Integral Product Operations Centers. And so, that whole software was built out of civic integrity. What else was out of as civil integrity? Digital oversee was out of civic integrity. That's like how do you get people to be safer online? The civic data teams are out of civic integrity. So, like, how do you know who was a civic actor, that sort of thing, or where is a polling place. That's all I could think of off the top of my head.

Q And you said at the beginning there were full pods. Did it change over time?

A It did. So I know there would be -- so I left civic misinformation when there was a reorg, so reorganization. So within Facebook, things that are frequently shuffled. And the -- I know the civic integrity team got changed. I think one of the measures of that change was the entire rapid response team got broken out into, like, their own pod. Like, my team did not have adequate support at the size that it was. So, like, we didn't report to an engineering manager who was appropriate for our problem space. And my team got moved under an engineering manager who was appropriate. And, unfortunately, now there was two product managers which was why I found a new project. Yeah.

Q Okay. And when did this change happen?

A It happened in February 2020.

Q And was there a change after that to the structure of civic integrity?

A I assume there was at some point, but I'm not aware of it.

Q Did civic integrity continue to exist at the time you left Facebook?

A It did not. So Facebook dissolved civic integrity in early December, like, maybe, the first week of December 2020. I have heard from reporters that they needed to do this before the 2020 election, and that they told it to me, but I did not know that firsthand. I, unquestionably, believe that dissolving the civic integrity played a role in the January 6th event, because I -- you know, it's kind of like in 9/11, like, there was -- there were a number of like key roles that were not staffed in the Bush White House, like, right after the transition happened. And as a result, like, there was just less, like, oversight going on. And one of the documents I have talks about like what levers were turned on at the time of the insurrection. Like things that were originally implemented for the November 2020 election, but were not turned on still by, like, 5:00 p.m. on January 6th. And I, unquestionably, think like part of why more practical things were not done because they resolved civic integrity when they did.

Q So you mentioned an analogy of certain key roles were not filled. Do you know which kind of key roles we are talking about, or is it a general feeling you have?

A When they resolved civic integrity, they went from having a single person who could make decisions in the play of resources who was focused exclusively on civic matters to having that team spread out across a huge number of stakeholders.

I don't know who would have had a similar level of, like, focus on things like social stability once that happened. It might have been as hot as, like, with any person who had a similar, like, sweep across all those problem areas now, whether it was Guy Rosen. I don't know that for sure, though. But I think if Samidh had still been in charge of civic integrity, there would have been a much greater focus on the evolving situation. And, like, I'm not sure who would have been that, like, singular responsible person once civic

integrity was resolved.

Q Like the evolving situation you mean?

A Like prior to 9/11 -- not 9/11. Prior to the insurrection, like, there were -- like it's not like we were surprised by Jan 6th. Like there was in the -- I have only -- I know from the documents, I don't know firsthand that there was like an awareness in things like stop the steal or voting. And the -- I think, like, given on how other situations like Ethiopia were handled, once civic integrity was still in place, I think there would have been much more proactive discussions on that if Samidh was still in charge.

Q And are you aware of any proactive discussions on that?

A Not -- I am not aware firsthand. I only know of things from the documents in this case.

Q Okay. And I think we might get to those documents during our interview today. So maybe we can come back to that. What got harder when civic integrity -- this is another way of maybe getting on this question -- what got harder when civic integrity was closed or its responsibilities shifted?

A So our agreeable (?) context in how like governing structure of how civic integrity works. And I don't know -- so by that point, I was focused mostly on the threat intelligence org. Prior to civic integrity being dissolved, there were meetings, I think, it was like twice a week where projects were reviewed, or, like, major high-level issues were, like, flagged and discussed. You know, like evolving situations in Ethiopia, that kind of thing. And like major -- and projects. So if they had staff that they had done projects, like, prior through and then like when projects were launched. And I don't know what equivalent meeting existed in that space once civic integrity was dissolved. Right, because, you know, Facebook likes to describe it as civic integrity was so important

that they integrated into other parts of the company. But the reality is that when you take, you know, people who are comprehensively looking at these kinds of problems, and you distribute them across different teams to have like slightly different priorities, you end up having less robust clarity into what is going on or what could be done to resolve it. I am not sure what that equivalent mean was afterwards.

I don't know if there was a question around empowerment, but like, you know, on why those issues -- they're not like single-thrust issues. It's not like, oh, there is problem with pages, there is a problem with groups. You know, there is a lot of these functional teams that focus on very small slices of the problem. Like civic integrity was intended and was like a pretty cross-cutting view on how Facebook was working, especially in a civic space. And I don't know that the equivalent mechanism for that was once civic integrity was resolved.

Q So you said that it was intended to be cross-cutting. What problem makes that?

A Just like the sheer diversity of problem spaces that we worked on, right. You know, like, in order to do social cohesion, for example, you work on hate speech classifiers, or like violence inside classifiers. So classifier is a system where a computer says is this thing X, or is it Y, right? And, you know, you do that, but you also do things like you analyze what kind of content gets the most distribution. You analyze what are the mechanisms of how Facebook is seen -- used in places that are unstable versus places that are stable. Like, you look at the problem from there more holistically. And the way a lot of other teams are designed on Facebook is they're all designed around having very, very criticals that are measurable. And the thing about evolving situations is that evolving situations usually are not clean measurable. And that's like an organizational flaw in terms of, like, Facebook, especially post November 2020, and U.S. 2020 elections

was in an evolving situation.

And when we have something that was that dynamic, you don't have time for the kinds of one cycles that are done on Facebook where you might have like a researcher develop a metric, and then later on, like, someone forgot assertion to that, like, oh, we're going to use that metric. Like that doesn't work when a situation is dynamic enough.

Q So Samidh Chakrabarti, which I know this last name is spelled C-h-a-k-r-a-b-a-r-t-i?

A That sounds right, yeah.

Q So he was the head of civic integrity. And civic integrity when was split up, there was no one person who was overseeing that?

A Because we were announced spread across the company, like spread across integrity, so people were looking in different spots. And there was like a much, much smaller node that was within ecosystems. So like there is new team that was formed like ecosystems. And like the person who was -- God, what's his last name? It's not Robbie Iyer Kashkik. Kashkik Iyer was the director of engineering for civic integrity, and so he went to ecosystems along with --

Q Could you please spell Kashkik Iyer?

A So Kashkik is spelled, I think, K-a-s-k -- gosh. Maybe K-a-s-h-k-i-k, Kashkik? Something like that. And then Iyer is I-y-e-r.

Q Okay. Thank you. Sorry. You were saying?

A So Kashkik Iyer was the head of engineering for civic integrity, and he got moved with a small slice of engineers to those systems. But from the product side, like we described before, there are engineering managers who kind of manage, it looks like, the health of the engineering org. And there are project managers who helped define, like, what is the problem we're going to solve, or problems we're going to solve, and like

what is the strategy in the form of work to solve those problems.

That is unusual to identify the problem and like how you would solve it, or like triage amongst decisions raised by his reports. And the role for Kashkik was like he managed the health of the engineers within his org. Like -- and, so, moving Kashkik to a place and not having someone like Samidh have the equivalent authority, like I don't know who played that role for him once he left. Like I don't know who got backfilled in specs. I'm not aware of it.

Q So civic integrity -- was Samidh an advocate for his -- his, some more of his ideas in the latter company?

A He was definitely an advocate for it. He was pretty well-connected. He had been with the company for a number of years. I think by the time January 6th happened, he had been there for at least 5 years, if not 6, which is a long time at Facebook. I think people who don't work in tech don't understand like spending 2 years at a company is actually a long time. Because your salary goes up when you change companies, and, so, people move between places much more frequently than people do in the general economy.

So Samidh being there for like 6 years is really quite a long time. And he clearly had been through enough experiences at Facebook like what happened after 2016. So that's like things like the information operations from Russia where he felt -- like he felt like that's secure enough, and he felt, like, adamant enough that he would push quite strongly when there were problems, because he had seen what damage Facebook could do.

And when civic integrity was reorged, I don't know who played anywhere near that, like, level like a role at that level of clarity, you know, like having both the influence and the willingness to stand up for those issues.

Q Do you know who decided to reorg civic integrity?

A I don't know. I assume Guy Rosen was involved, but I don't know beyond that.

Q And do you know what reasons were given to reorg civic integrity?

A They claimed civic integrity was so important they needed to integrate it to the rest of the company. Lots of integrity. Like they needed to cross-pollinate other parts of integrity with civic integrity. But Facebook -- that's not really how reorgs work. So reorgs happen because, like, Facebook claims it is because, you know, it's so dynamic there is a need to frequently reorganize the company. But the pattern that you see over and over again is that, you know, some teams are viewed as having like a better strategy than other teams. And, so, when you have a situation like that, there is a team that -- there is a team that more of the authority in the leadership direction when that reorganization happens. And let's just say civic integrity was not the organization eating another team. We were eaten by the larger community integrity team.

I think the other thing to keep in mind is when -- let me see how I describe it. So I think I knew what's happening as early as November 2019, also working on narrow cast misinformation. So within civic misinformation, we worked for a semester -- a quarter on something called narrow cast misinformation, and we developed a system for sedimentation of population. So, all integrity problems on Facebook happen in a nonevenly distributed way. So they often follow, like, power laws, where, like, the difference between someone who is, like, at a 95th percentile exposed to a problem and 99th were just huge differences. And between 99th percentile and 99.9th percentile, it's another huge chunk worse. And we developed a system for being able to break the population into small populations to understand how that overexposure happened. And I remember when we went through that project, we sat up front before we began.

You are asking us to do a million possible things in the timeframe we're asking for it, like 12 weeks. We will not do measurement this quarter. We just can't. Like we can't do measurements in this quarter. And partway through the quarter, Samidh started putting a huge amount of pressure on us for measurement. And the reason for it, at least as he described it, was that the larger community integrity team not think the problems that civic integrity worked on were sufficiently measured. Because success in Facebook is defined by, you know, having a way to measure your problem in the meeting that measures over time. Like, I think Samidh must have been put under a huge amount of pressure because he put our team under a huge amount of pressure, even though I am not going to remotely tell [inaudible], I don't think you actually care. Narrow cast misinformation is extremely hard to measure in the context of how Facebook works. And it took probably -- I am not kidding here -- on the order of 3 whole hours for a team of four people to explain to Samidh why it was we couldn't measure our problems.

So the fact he was pushing us so hard shows you how much pressure he was under as the problem was measured. And so, I think it happened with regard to civic integrity for his community integrity was there was a perception for quite a while that we were an underperforming team. Because like civic integrity was working on these large, messy ambiguous problems. I think we were doing things that were actually moving the needle in those problems, but we weren't doing it in a way that was aligned with how Facebook operationally worked. And so it was not viewed, I think, as being as successful as, I think, we were, but that's just context.

Q Okay. Well, then I think we will get to measurement in a minute. I guess just a fine point on it. Did you believe there was another reason for spreading up civic integrity than the one given?

A I think there was also a thing that civic integrity uncovered problems. It

was, like, once you know about a problem, you know, then one could say you have an obligation to fix that problem if you cause that problem, right? Like, civic integrity was home to a lot of researchers who did really good research on how is Facebook impacting the world. And I think it was viewed as, like, opening cans of worms that were difficult to close. And because Facebook didn't want to invest the resources, or couldn't because they had trouble hiring on, like -- I think there was a view that like civic integrity did not add as much value as it added risk.

Q And do you think civic integrity should have gone split up in December?

A I don't think so. I think there are things that probably could have been done to help the team be happier and more functional, but I think splitting it up is not the solution.

Q Now, we're going to go through some quick factual things just to lay some foundation for some of the documents you shared with us.

A That's good.

Q So you decided to -- okay, let's ask a specific question. When did you decide to take the documents with you when you left?

A I began thinking about it for a long time over the course of the summer of 2020. It took me a long time to come to the conclusion that I -- that I would need to do that. I could see there was sufficient managing problems that I -- I was struggling with what was my responsibility in that situation?

When civic integrity was dissolved, I felt a pretty high level of conviction that there was no chance that the problems that I was most worried on, which were specifically on ethnic violence in lesser-developed countries. I had no faith that that was going to be resolved in the structure. And there is a difference between when I became convinced that I would need to blow the whistle, and when I began to really accumulate content.

And, so, I started accumulating a little of it in December of 2020. But I would say I got the vast majority of what I captured in the last, maybe 2 months that I was at the company.

Q Got it. And how would you go about offering these documents?

A I used a cell phone, and I took pictures in my laptop screen.

Q And the photos that you shared with us are accurate representations of the contents of the original documents?

A Yes.

Q So all of the documents you handed over to the committee used the same process of FOTEMP of taking photos of them on your computer screen?

A I think there are two phones I used. A few of them were taken using my personal cell phone, but the vast majority were taken using, like, a cheaper phone. And part of that is that my new phone is -- it takes much larger photos, and it was just a data management problem. That there were certain instances, like, on January 6th when I am pretty sure I used my personal phone, just because it was web-based and readily available.

Q And did you alter any of the underlying documents before you took photos of them?

A No.

Q Did you see any kinds of documents or information that were of interest to you that you were not able to capture by taking photographs of them?

A There are maybe -- maybe on the way of two or things that I really wanted, but Facebook had already deleted them. Facebook has a pattern of deleting things inside of the company that they think are dangerous to the company.

Q Do you recall what those were?

A Sure. So one -- the one that I most wanted that I couldn't bring with me was -- when I was working in civic misinformation, I had a data scientist named Wen Xie. W-e-n, I think X-i-e. I could look it up. And she worked on civic ranking after she worked on civic misinformation. And all of her notebooks, so when data is extracted -- when data is extracted -- talking for an hour straight does stuff -- things to your voice.

Q You can have a sip of water if you need it.

A When I was trying to capture things, or like in the process of capturing things, there was an analysis that she did around concentration of voice. So that was a question of how few people control how much of the voice, like, reach on Facebook. And the United States is the most worldly democratic country for voice. And the 12 percent of the population makes up 80 percent of all the voice. Many countries in the world like on the order of 1 percent of people make up 80 percent of all the voice, because Facebook's system hyper amplify.

And I went to go find that notebook and discovered that all of her notebooks had been deleted. Like all of them. And most people who leave the company, they don't delete all their notebooks, because there is a central serving called ANP at the time. And it may have changed having a different name. But ANP was like the notebook server.

And I found a screen shot of the notebook I wanted in our chat logs. But even within our chat logs, they had purged, like, photos and things related to her. And so, I was only able to get a thumbnail. I couldn't get the original even full image. And I saw that delineated content again in another collaborator, but we can go into that later.

Q Is there a reason you don't want to go into it now?

A Not now. It's just like it's pseudo-related. It's not directly related.

Q Oh, okay.

A It centers around fake accounts being generated in like countries where Facebook had its subsidized internet programs, and those accounts are being used for abuse of services across, like, abusive acts across Facebook.

Q Right. I think that's -- we can come to that at a different time. So there were photos you shared with us last couple of factual questions. There were photos you shared with us, were any of those taken by anyone else on your behalf?

A No.

Q Now, we are going to turn a little bit to Facebook's corporate structure. So background to bigger questions, what other things do you think is most important for the committee to understand in Facebook's role and what happened on January 6th, if it had a role?

A Yeah. Give me one second. What is Facebook's role on January 6th? So in organization structure, how did that impact? So Facebook is an extremely flat organization. So it focuses on -- it has this kind of utopian idea that if you pick the right metrics that you -- if you pick the right metrics, you don't have to tell people what to do to solve the problems. That people can act independently. And as long as they move those metrics, it's okay.

But this is an okay management philosophy if the metrics themselves are not the problem, right? And in the case of Facebook, Facebook made changes in 2018 to how its algorithms worked. It ended up prioritizing, polarizing in divisive content. And people began identifying almost immediately that the metrics were driving these problems. And I don't know this next part for sure, but I think the reason why a number of problems -- I think some of which we're going to discuss a little later on happened is because in order to fit it, to make the system overall safer, why it's so like prioritizing less

extreme content, and having safer ways of running groups, all of these things, those actions would also end up hitting core metrics and would likely then also hit people's compensation. Like if the metrics or the problem, that becomes a really big challenge for fixing it, because, you know, if people have made lots and lots of choices around, like, what they focus their time on, what they built in order to move those metrics, it's very disruptive to change the metrics.

Anything else with the organization structure? I think the fact that Mark has so much power is problematic, because it means that whatever his philosophies or opinions are have outweighted significance because --

Q You mean Mark Zuckerberg?

A Mark Zuckerberg, yeah. Because, you know, the people who got promoted, the people who -- actually, this is a good organizational thing. The Wall Street Journal did a very good analysis on how the people who were in charge of integrity, the managers and stuff, largely came from the growth team. If people were not promoted internally, they were promoted from outside the team. And they were, specifically, promoted from a private team that was at the company that was focused on making the company bigger, not making the company safer.

Q So, and how do you link that to January 6th?

A There were lots and lots of options that Facebook had in making the overall system safer at lots and lots of points in time, but those things slowed growth down by little slices each time. When you have a bunch of looters (ph) in the safety org who, you know, always found the places they were trained up in the company's growth, if growth complains about an action, they're going to be less willing to take a harder line on it.

And so, that's how it is to Jan 6th. There is many things that could have been done to make the factors that flamed Jan 6th, like the stop the steal movement stuff.

There is many things that were content neutral, like they weren't -- they weren't about picking good or bad or ugliest. They were about making the system less twitchy, reactive, viral. But all those actions also would have made the company grow a little slower. And so, having these in place you would think that growth is of such painful importance, that's problematic.

Q So his decisions weren't necessarily made on January 6th?

A No, they were not made on January 6th. They were made over months and over months, even before, even before the 2020 election. Yeah, go ahead.

Q Could you elaborate on specific critical decisions, any inflection points?

A Any singular ones. I did not work on that part of the company at that time. Like, I was working mainly on counterespionage. Actually, it's not true. So downstream MSI. So MSI is meaningful social interactions. Downstream MSI is a pretty critical factor in how content is prioritized at Facebook. So you could imagine --

Q Does this relate to an earlier reference to changes in the importance of metrics?

A Yes.

Q Okay.

A So meaningful social interactions is the change that was made in 2018. So it used to be that Facebook prioritized -- the bare measure of success was, could they keep you on Facebook longer? Like the longer you stayed on Facebook, the better whatever -- like, if you were saying is A better or B better? And if you stay on Facebook longer, that was a better -- thank you. Coffee delivery.

The -- in 2018, Facebook made a change to what its definition of success was to prioritize things called meaningful social interactions. And the meaningful part was, basically marketing, because social interaction to things like likes, comments, reshares,

messages, that kind of thing -- likes, comments, reshares, and -- so you can imagine for -- but at the same time, if that comment was hate speech or the comment was bullying, it's still counted as a meaningful social interaction.

And they made a change in November 2020 where if there is bullying and hate speech in a comment, it wasn't considered meaningful. But the things remembered there is only a tiny, tiny fraction of hate speech comments or bullying are caught. In effect, you still have things like bullying and hate speech counting as meaningful social interactions.

So now, let's imagine two different ways prioritizing content within the news feed. So you could say, Hey, we're only going to look at, would Frances react to this content? Or we could say, if Frances reshared this content, you know, her friends might like, comment, reshare on it. And so we should really word the content for the chance that her friends will interact with it as well. That prioritizing based on the downstream effects, so not my -- not my meaningful social interactions beyond the reshare, but on all those other people's reactions, Facebook identified very early on that that was an extremely dangerous factor. And by early on, I mean 2019. And a number of people tried very hard to remove this factor from the ranking because people don't like reshares; like they don't interact with them as much. There is a really beautiful graph in the cache that shows how much you like original content from your friends. And you would like original content from your 20 closest friends.

You only like reshares from your five closest friends. So the bar for eliciting a reaction on content that is reshared is much, much higher. It happens to be the content that disproportionately can trigger a reaction when a downstream manner is the most extreme content. And because removing this factor would hit the core MSI factor, there are multiple, multiple moments where people try to get this out of the ranking because it

1 was over prioritizing extreme polarizing divisive content. And because it also ended up
2 decreasing the core metric, it was consistently not done.

[12:07 p.m.]

BY [redacted]

Q So I have a couple of followup questions to clarify some things.

Is MSI a sum or a proportion or something else?

A I think it's a sum. So, like, it's a weighted sum. So different kinds of actions have different values. And there are copies of those weights at different points in time within the document cache. So you can see the importance of different factors at different points in time.

Q And so you said "core MSI," and what does that mean?

A The core MSI is, you go and you count all those interactions that happen on the site, and you add them up; all those together, that is MSI.

Downstream MSI is the fraction of the MSI that comes not from immediate people but from the next top down or further.

Q Okay. And MSI was something that people tried to maximize?

A Yes. I did not have a bonus that was tied to MSI, but I guarantee you other people had bonuses that were tied to MSI.

Q And the change that was proposed relating to downstream MSI, was this a change to what counted as MSI, or was it a change that was expected to decrease MSI, or was it something else?

A So downstream MSI was a ranking factor. So there's many different things in the scoring that caused content to, you know, get more distribution or less distribution. And the --

Q And by "distribution," you mean appear higher on the timeline?

A So appearing higher on the timeline would give you more distribution. So,

for example, like, distribution is just, like, how many times did people see it, how many people saw it, that kind of thing.

And so downstream MSI -- so MSI is, like, basically an aggregation of, like, you know -- every "like" is worth, I don't know, one, and, like, every reshare is worth some much larger number because they're rarer, that kind of thing. Downstream MSI is just -- so, like, let's imagine the experiment where they decided to add downstream MSI to the ranking.

So they would've had a version of MSI that did not consider downstream effects. And they would've measured, what's the total amount of these meaningful social interactions that happens on the news feed? And then they would've had a second version that included this extra ranking parameter, and they would've accounted, across all the people who interact with the news feed, how many of these meaningful social interactions happened once that ranking factor came in. And because the aggregate number of interactions went up, it was considered a better thing than the previous one.

Q We may return to MSI, because it sounds like that was an important thing to talk about. But I'm going to take a step back to talk a little bit more --

A Actually, I want to take one moment. It's extremely important to talk about, because there is a document called "Mark notes from soft interventions review" or something. That document is extremely important, because people suggested we should turn off downstream MSI in at-risk countries because we know it's dangerous, and Mark said, if it hits core MSI, we're not going broad with it.

Q Okay. Let's speed through some of these things, so we can get to some of these documents you've been referencing.

A Yes.

Q So, in terms of understanding how Facebook works, I'm going to name a

couple people; please just let me know if you think their decision-making or responsibilities or actions is relevant to something we're looking at, based on what you saw from inside the company.

So, Javier Olivan?

A Oh, okay. So Javier --

Q We can just do this pretty quick.

A Actually, no, I don't know Javier. I was thinking of Henri (ph). Never mind.

Q And Naomi Gleit?

A I know that she existed. I don't know what her role is.

Q And we mentioned Guy Rosen. Joel Kaplan?

A Joel Kaplan -- I don't know exactly what he did.

Q Okay. John Hegeman?

A Unquestionably had some role.

Q In what?

A So he was the head of ranking for the news feed.

Q Okay.

A I don't know -- I don't have, like, a smoking gun on John, but given that he was the head of the news feed for, like, those prioritization functions, I guarantee you -- like, actually, I'll give you an example.

There's really interesting documentation in the cache on this. People tried to get anger, like, valued less in the news feed's ranking. So, like, remember I talked before about the idea that MSI is a weighted aggregate? They wanted to change the weights on the different emotions, because there was lots and lots of research that it's easier to inspire people to anger than other emotions. Okay? And they almost changed that

ranking at the end of 2020, and at the last moment it got blocked.

And I can't imagine that John wasn't involved in some way with that -- with that reversion. And I encourage you to go read the comments on those, and I can try to find those specific documents for you afterwards.

Q I think I know what you're referencing. They may have been reported in The Wall Street Journal and others.

So Justin Osofsky?

A I don't know who that is.

Q Okay. And then Brian Fishman?

A That name sounds really familiar. I don't know what he did.

Q Okay. And Sheryl Sandberg?

A She probably had some role with Strategic Response, because that was under her.

Q And Molly --

A And I also know there were other situations where there were policy changes suggested by Civic Integrity where she would've been, like, the initial decider.

Q Okay. Molly Cutler?

A Name sounds familiar. Don't know what she did.

Q All right. And then Mike Schroepfer?

That's -- oh, I'm sorry. I haven't been spelling these for the court reporter. S-c-h-r-o-e-p-f-e-r.

A Yeah. I know that name. I don't know exactly what he did. I saw his name on the docs.

Q And you mentioned that you don't have a smoking gun on John Hegeman. Is there anyone -- that's H-e-g-e-m-a-n. Is there anyone who you do think that the select

committee should look into?

A Yeah, I mean, like, the soft intervention stock, I think, is a really big problem -- is a really big problem. Right? Like, I contributed to that working group. Because they came in with a lot of options that would've made both the 2020 election safer and, if they had been left on, would've made the run-up to the inauguration substantially safer.

And the fact that, when Mark reviewed that document, a lot of people who are extreme experts in this field -- right? Like, these are the best people in the world -- said, hey, we consulted 60-plus experts inside the company; this is what we came up with.

And the fact that all those things weren't done, I think, is a huge problem. Because these were not things -- like, Facebook says, we're against censorship. These were not things about censorship. These were things about how the product was designed. And the fact that growth and profit was prioritized over safety, I think, is a huge problem.

Q And --

A And that directly ties it to Mark. That's why that document is so important.

Q Okay. That was, I guess, the next thing I was going to get to, and then we should get to some of these documents.

A Uh-huh.

Q So you mentioned there was a flat organization, and you mentioned that Mark sometimes made decisions on these kind of things. Can you talk a little bit about how decisions were elevated and how people knew they were authorized to make decisions and when they elevated them?

A Hmm. I don't know the exact heuristic on when things went to Mark. I think when there were conflicts across groups -- so, for example, often, to do something

that involves safety, it trades off with growth, and I think when the impact was going to be large enough, it would go to Mark.

When it was viewed as, like, needing to have a mandate -- so, like, you know, if you are going to do something like the soft interventions actions, those touch lots of different points in the company, right? And so having a mandate saying, like, "Hey, we need to fix this problem" was important in that kind of situation, because you're asking a lot of different people with different goals to do things.

Beyond that, I don't know.

Q All right.

So you mentioned that sometimes things went to Mark when there was a conflict between teams. Were there any processes where Civic Integrity had an idea and another team regularly disagreed?

A So I think there's an important thing to keep in mind, which is -- so, like, remember how I mentioned before, there was, like, a review called "civic launch review"? So that's the meeting that happened, like, twice a week. And it wasn't really a launch review; it was also, like, a status update kind of thing.

There was definitely -- and, remember, organizations self-censor over time. Like, if you learn that if there is a tradeoff between this metric and this metric that it's just not going to go anywhere, you end up killing things at earlier stages; they don't go to a final review. Right? And there are documents in the cache that describe this idea of, like, what is an acceptable loss in, say, the number of sessions people do -- so that's, like, the number of times people open the app per day -- and, say, 10 percent more misinformation.

And I definitely watched projects go to civic launch review where something might look promising but, if it hit, you know, some form of sessions, some form of MSI, that kind

of thing, that it would just not get further than that, because it was viewed as, like, a nonstarter.

And I'm sure there are many things that never even made it to launch review because people, like, you know, showed it to their manager and their manager's like, oh, this hits growth in this way or that way; there's no point in investing further in it. Right?

Q So a lot of good ideas didn't make it --

A Yeah. Like, that's probably what's so -- it's interesting. I was contacted by Katie Harbath sometime after I did my disclosures, and she said that -- so one of the things you could do, right, is you could cut researchings at two, like, two hops. And so the idea behind that is, like, let's say I write something -- or let's say Alice writes something, Bob reshares it, Carol reshares it, now it lands in Dan's news feed. So there's been two hops there. Imagine if Bob -- not Bob -- Dan now needed to copy/paste that content in order to share it further. That is cutting researchings at two.

And Katie Harbath messaged me and said, like, "We only ever discussed that in launch review once," like, basically telling me, "This is irrelevant; we only ever discussed this in launch review once." Because it has a huge impact on misinformation, right? It's like a similar size impact on misinformation as, like, the entire third-party fact-checking program, right? And she was like, "This is irrelevant. We only discussed this in launch review once."

And the part that frustrated me so much about this comment is, of course you only discussed it once. Like, it hits growth metrics. Right? There's places in the world where 85 -- not 85 -- 35 percent, 3-5 percent, of all the content viewed in the news feed comes from reshares. So I'm sure, like, in that case, like, they looked at the metrics and said, "Oh, it decreases the total amount of content viewed by 1 percent," or something; "we can't do it."

And so it's those kinds of things that are so frustrating to me, because, like, there was such a presumption that if growth was harmed basically in any way, you just couldn't do things.

And to make it worse, there were some changes where it didn't hurt the number of sessions, it didn't hurt total content viewed, but it did hit MSI. And, even then, those changes would not get shipped, because I think people's bonuses were tied to MSI.

Q Okay.

A And that's what's so frustrating, because that's a thing where it's not even about the economics of the business. Right? It's not like the business was harmed in any way. They didn't get any fewer ad dollars. But because organizationally it'd be difficult to do something that would harm MSI, it wasn't done.

Q And speaking of data, have you heard of a concept called "the coefficient" or "the coefficient of the graph" at Facebook?

A Is that in the connectedness? Is that, like, how many hops it is to people, yeah?

Q You're the witness. You tell me.

A Oh. I think -- there's a lot -- we have a -- I don't know if you guys were given the glossary along with the docs. We have a glossary that's, like, hundreds and hundreds of terms long. So, like, I'm doing the best I can, because there are a lot of coefficients and things.

If coefficient is, like, the connectedness of the graph, that means, like, how many hops does it take for any given person to reach any other person in the graph? And so, like, if you had a more densely connected graph, it would take fewer individual hops. And, like, the concept of six degrees of separation is the idea that any person can reach any other person in the world through six hops. I think the connectedness of the

Facebook graph is, like, three and a half or something. But I'm not sure.

Q So I think we're going to move to --

A Actually, to hit on that, a really important detail on that is, like, why would the connectedness of the Facebook graph be denser than the real-world graph, right? And the reason for that is that Facebook allows super-connecters to exist much more functionally. Like, you know, a person that's connected to thousands and thousands of people. And when you have amplification systems like that, you end up having a small number of people change the dynamics of the overall system substantially.

Q Great. Okay.

So I think we're going to start getting ready to talk about some documents now. Are we ready to turn to that?

A Uh-huh.

Q Lay a little bit of information down before we get there. So --

A Just give me 1 second.

Q Okay. That's fine.

So break-the-glass plans at Facebook, are you familiar with those?

A Yes.

Q What are break-the-glass plans?

A So, like I mentioned earlier about the idea that Facebook doesn't -- like, Facebook knows lots and lots of ways to make the system safer. You know, part of the reason why the cache of documents is so large is I wanted to make sure the history saw that Facebook knew ways of reducing misinformation, reducing violence-inciting content, all these things, and that they chose not to use them.

And Facebook had a process for turning those things on when things got bad enough. So the idea of "break the glass" is the idea of -- it's "break glass in case of

emergency." You know, it's like a fire detector -- fire alarm.

And I mentioned at the beginning the idea that one of the projects that we were pursuing initially was the ability to quantitatively describe how bad a situation was. Because the only way Facebook could, within their policy context, be able to change the dynamics of a system was if they could rationalize in a, quote, "objective" way that the situation was materially different than the baseline operation of the site, and, therefore, they could change -- you know, actually turn on these safety systems.

So, in the case of January 6th, there was a break-the-glass plan. But, given the photos -- I believe you guys have received the photos that are --

Q So maybe at this point we should put document 3 on the board and mark it as exhibit A.

A Okay. How do I see it?

Q It should be popping up.

A Okay.

Q We'll give it a minute.

Do you see it?

A Yeah. Perfect. One second.

Q So we'll call this --

A Actually, before we dive into this one, you should've received a second one that has a timestamp.

Q Yeah.

So can we go to document 4 and mark that exhibit B, please?

A Yeah. Yes, that's the one.

Q So you recognize both of these documents?

A Yes.

Q And you took these photographs?

A I did.

Q And you shared them with the committee?

A And, for your context, on the timestamp --

Q Wait. Was that a "yes"? Sorry.

A I'm sorry.

Q You shared them with the committee?

A I took those photos, and I shared them with the committee.

Q Right. Thank you. You were going to say?

A And I wanted to make sure it's documented for the record that the 2:00 p.m., like, the 1:59 p.m. in upper right-hand corner of document 4 refers to a West Coast time zone. So I was in -- I was near Grass Valley or Nevada City, I think more close to Nevada City, California, when I took this photo.

Q So this is around 5:00 p.m. eastern time, or 4:59 p.m. eastern time?

A On January 6th, yes.

Q On January 6th.

A So the fact that the timestamps on this thing are what they are, it makes me believe that this plan may have either been created at the last moment or was created -- like, may have been created even the day of January 6th. I just don't know.

Q Okay. And the document 3 is the same picture as document 4 just cropped, correct?

A Let me -- can you toggle between them? I think it is.

Q There's another one, which is document 5 --

A And then you go to -- can you go to 4 again? Yeah, I think they're the same one.

Q Okay. Great.

And what does the color green mean to you in the column E?

A Uh-huh. So, to walk through the different columns, right now, the ones that are of most importance are the name, which is the description of the intervention; the description, which says what does that name mean; the owner was, like, the stakeholder who is responsible for, like, actually making sure that gets turned on. They might not be responsible for the final decision to turn it on, but they're, like, responsible for -- like, maybe they built that system or they are the ones who were in talks with engineers to actually get it actually ramped up.

The criteria to recommend each break-the-glass intervention, there is something -- there is a catalog of break-the-glass interventions. So there was a tool that specifically -- I think these things are called EPICs, E-P-I-K -- excuse me -- E-P-I-C. I don't know what the acronym stands for. But there is, like, a catalog of these interventions, and each one has a criteria on when can you turn it on. And there might be a possibility of having different criteria per situation, but I'm not sure on that one.

And then the status column is -- so green, I think, means, like, the criteria had been met. And then status is whether or not that thing is on yet.

And then the 1:00 p.m. notes, I believe, are for 1:00 p.m. of that day. So, at this point, it's 2:00 p.m. So the meeting, I think, had just finished. And they would've updated this document live in that meeting, just for people's awareness. The 1:00 p.m. notes are basically the discussion of, you know, are we turning this on and where is the status on it.

Q Do you know who decides to develop these EPICs or break-the-glass measures?

A So, in the run-up to -- so, in Q4 of 2019, there was something called a

lockdown. So lockdowns happen very, very infrequently --

Q I'm sorry. One thing. Are we talking -- is Q4 at Facebook in the way you're describing it the last 3 months of the year, the calendar year?

A Yes.

Q Okay. Thank you.

A So, from October through December 2019, there was something called a lockdown inside of Facebook. There have been very few lockdowns in the history of the company, I think on the order of, like, maybe six or seven. So it's things like: Facebook was not mobile and they did a lockdown and built up mobile apps; or Facebook was on iPhone but it wasn't on Android, so they did an Android lockdown. Like, there was a very small number of these events that were considered of high enough importance or, like, existential risk to the company that a lockdown occurred.

In September of 2019, there was a review -- like, they did a red team exercise where they basically said, how could the U.S. 2020 election go off the rails? It's 12 months out. How could it go off the rails? And they came to the conclusion that there were so many problems -- and there's a document in the cache on this -- there were so many problems that, when they made a grid of the surface areas and the problem types and they rated them green, yellow, red, there was so much red that they had to have two colors of red to differentiate between what they were going to try to fix and what they were not -- they were just not going to focus on.

During that lockdown, there were a variety of different teams responsible for different parts of that problem space. And part of the responsibilities for some of those problem spaces was developing playbooks -- like, they were really called "playbooks" -- for, when this problem is encountered, how will you deal with it? Like, how will you detect it? What will be the escalation path? What are the

break-the-glass measures or the EPICs that you're going to use in order to do that?

And so each one of these interventions would've been developed by someone. And I think what RE45 or PE1 or those kinds of things, I think those relate to the catalog, like, catalog numbers for those EPICs. So different people would've been in charge of, like, you know: I work on violence-inciting content. Here is the playbook and the tools that are available to deal with violence-inciting content.

Does that make sense?

Q Yes.

Okay. And are you familiar with "break the glass" relating to the election?

A I don't have that documented. But the thing that is interesting about this specific document, the "Capitol Protest BTG Response" document, is that there are multiple sections within that document, some of which are new interventions, that could be turned on for January 6th. And some are under the heading "U.S. 2020 Levers Previously Rolled Back," so these are things that were in place for 2020, for the November 2020 election, that were then turned off afterwards.

Facebook has since defended these actions as saying that they are pro-free-speech. Lots of these interventions did not involve free speech. It involved choices in terms of what Facebook was giving more distribution to, like live video, not censoring which live videos are okay or not okay.

Q And these were all on January 6th at 5:00 p.m. was what you believed this document was created for --

A Yes. I believe this document represented the state of the world at 5:00 p.m. eastern on January 6th.

Q But there was a period between the election and January 6th where Facebook was taking things down --

A Yes.

Q -- in terms of --

A I don't know how fast they took things down. So it's a thing of, like, I think, you know, if you subpoenaed Samidh Chakrabarti and you said, "Hey, when did they turn these things off?", like, he would have a much better understanding of, like, well, how quickly did they turn off these safety systems.

I think, given when they dissolved Civic Integrity, I'm guessing they had turned off those systems by then. Because I think part of why they got rid of Civic Integrity was they viewed themselves as succeeding in their mission, i.e., nothing, like, blew up on election day, therefore they were done. Right? So I assume they [inaudible] relatively rapidly after it became apparent there wasn't going to be [inaudible] on January --

Q Wait. One second. One second. I was going to say, I think you were breaking up.

A But what was the question?

Q -- or could you please repeat it? It was -- you were talking about when break the glass came off.

A Oh, sure. I was saying that, given when they dissolved Civic Integrity, like, part of -- I don't know this for sure, but the perceptions of the team was that Facebook viewed itself as having been successful in the U.S. November 2020 election, because, like, there was no rioting or, like, no violence at the November U.S. 2020 election, and that, you know, they had succeeded.

And so I assume they turned off a bunch of those levers relatively rapidly after the November -- I think it was, like, November 8th was when the vote was.

Q And you were -- oh, I'm sorry. I had a little feedback. So you were working at Facebook after the election. And what was the feeling there? Was it a

general feeling of success?

A You know, I don't know -- but by that point I was working on counterespionage, and I was not as tuned in to the general feeling amongst Civic Integrity at that point. I think there was satisfaction that there wasn't rioting. I don't know beyond that, though.

Q So, going back to the document, you mentioned the table, starting on page 1, that's titled "U.S. 2020 Levers Previously Rolled Back."

A Uh-huh.

Q You referenced the one on page 4, if we can go to page 4, please, that's titled "New Opportunities."

A Yep.

Q I guess -- is that up there, or is it the next page or the previous page? I don't see the table title. Try the next page. Try two more down.

So there's that, which says, "New Opportunities."

And then on the next page, page 5, there's a table titled "Anticipate Growing Risk," phase 2. Do you know what this table -- can you help us understand what this table says?

A Where does it say "Anticipate Growing Risk"? I see "Follow-Up Action, Phase 2." Is there another table beyond it?

Q Are we in document -- are we in appendix -- I'm sorry. Are we in exhibit A, document 3?

[redacted] Let me just scroll back and make sure that it's the right one.

[redacted] Okay.

[redacted] This one?

Ms. Haugen. That one. Okay.

[redacted] Yes.

Ms. Haugen. I guess they wrote "Phase 2" twice or maybe -- yeah. I don't know what that is.

Okay. So what was the question?

BY [redacted]

Q What is this table? What is -- what is this table? Can you help us understand it, what this table includes?

You've mentioned the table above it, which is "New Opportunities." What does this --

A Oh, interesting. This is a really interesting document. The -- so let's walk through some things.

So, "Anticipate Growing Risk." So, if you look at the second row, PE -- actually, let's go through a couple of these. So let's go through RE34. That's row 2. So, live video. So "LVEQ" I'm guessing is live video quality.

So -- wow. That's amazing. So, how to describe this?

So Facebook has described why they rolled back the things they rolled back as they were protecting freedom of speech. The first row is really interesting, in that live videos only can act as live videos if people see them when they are happening. And so, in order to make sure that they show up at the top of your news feed, this row says that, before the U.S. 2020 election, if there was a live video, it got a boost of 850 times its normal quality score, which was, I'm guessing, to increase the probability that it would end up at the top of your news feed, right?

So this said, hey, like, let's return -- during the U.S. 2020 election, we lowered that boost down to 83.5X, right? Because, you know, low-quality videos should not end up at the top of your feed, right?

And the thing that is odd about this table is there's not a status column, and so I don't know if these actions were taken. Like, they're described in the description field as having been done, but I don't know -- I don't know what the timing would've been on those.

Q Maybe let's go to document 5, which is where we may have been earlier.

A But if we can finish that document real quick --

Q Oh, sure.

A -- I want to flag a couple things on these specifically.

Q All right.

A So what I was really intrigued by that table is, if you look at the second row -- so this is PE21 -- it said, "We removed 10 product boosts for civic and health content to prevent it from going viral for reasons unrelated to user value."

So there are -- remember how we talked about before the idea that, when you rank things in the news feed, you can either prioritize for the happiness of the user, so value to the user, or you can boost that content because of, say, value to the network. Like, if I reshare it, then that content goes out to my network.

There might be other boosts that are done for reasons that are in the best interests of Facebook's business. And what this basically said was, hey, like, we looked through and said, like, these 10 different boosts were not happening just because it was good for the user; it was happening for other reasons, and it happened to be causing that to go viral more.

Give me just 2 secs to, like, scan this real quick.

Q Well, why don't we scroll up to the top, and you can --

A Sure.

Q -- go through all of them.

So let's go to the top of appendix A. And I was going to ask you -- or, I'm sorry, exhibit A.

A Yep.

Q And I was going to ask you to just take a look at these and let us know if anything stands out both as something that was useful to consider or that was important for us to know about, related to the violence on January 6th -- because, again, this document, we understand from you, was related to responses from Facebook on January 6th -- and how it might've been useful in the weeks in between the election and January 6th.

So those two things are what --

A Sure.

Q -- I'm asking you to flag for.

A So let's go -- we'll start at the top and move our way down.

So the first one is -- the reality is that lots of repetitive content happens on the system, right? So, you know, people will make very, very similar posts, or maybe there's, like, a meme going around or something. And the idea on the first one is that you can actually cluster that content and have a cluster get reviewed, and that ends up leaving it so that substantially more content gets assessed. So that looks like a thing where, like, they could've been able to action more content had they been doing those clusterings.

The second one is around distribution. So, instead of -- so there's content that is judged to be, yes, this violates our community standards. So this is, like, hate speech, graphic violence, violence incitement, as listed in the table. That threshold is very, very high, right? So, in order to prevent false positives, like, removing content when it's not bad, a very, very small fraction of bad content is actually actioned. So other documents

describe that as 3 to 5 percent of the content is actioned.

What this second one -- so this is row 3, PE1. This is saying: Hey, like, here's the deal. We know that a huge fraction of the content with a slightly lower score than what we currently use to take content off is bad. So, instead of making this a binary thing, where we either take it off or we don't take it off, let's, like, demote it proportionate to our confidence.

So that would end up showing substantially less violence incitement, hate speech, graphic violent content. Because now you're applying a demotion to a much larger number of pieces of content. And showing less violent content likely would cause less violence.

The next one, which is row 4, says, like: Hey, let's increase the max value for violence and incitement demotion to a higher number. So it goes from 50 to 80 percent. And that applies --

Q So --

A -- to the -- sorry. Go ahead.

Q So, taken together, how would these change the Facebook experience?

A It would mean a substantially -- so, across all these levers, they're trying to reduce virality, which is, like, the likelihood that content will, like, rapidly spread like wildfire through the system.

Given that there are many, many documents in the cache that describe how the most extreme content gets the most distribution, if these interventions that had been live on November -- like, in the November U.S. 2020 election, had they been on in the weeks leading up to January 6th, extreme content, violence-inciting content, and hate speech would've gotten substantially less reach on the platform.

For the one for row 6, so that's PE20, this is an example of another one where

you're not specifically censoring individual pieces of content. It's saying: Hey, if your group has accumulated a lot of violations, just require admins to approve each post. Right? Because just that alone will reduce the amount of content that gets distributed within those groups.

Can you scroll down?

Let's see. So things like freezing commenting -- so this is PE17, that's row 7. Comments are extremely dangerous on Facebook, or at least they were at that time -- I assume they still are -- because, as hard as it is to figure out what the bad content is on a post-by-post level, because comments are so much shorter, they're even harder to identify.

And so this basically said: Hey, given that commenting and people piling on, like, basically, people encouraging each other or, like, amping each other up even more, given that we know that, for some of these particularly inflammatory groups -- so, like, for example, PE20, the row above, described it as, like, there's 57 fast-moving new groups that were particularly bad. Like, inside of groups like that, like, if you freeze commenting, you end up having less of that echo cycle, like, that less of people reducing themselves up.

Q And was there anything stopping Facebook from having these turned on in November?

A Nope. So they did have them turned on in November. The question is turning them back off, right, in, like --

Q I'm sorry. I meant after the election in November.

A So the only thing that would've kept them from having them on afterwards is that some of these actions -- so, for example, the live video boost, like, if you had that live video boost off or, like, if you weren't boosting it as much, so you're only boosting it

83 times instead of 850 times, you end up in a situation where, like, the live video product would grow less. You know, fewer people would encounter a live video at the time that it was on.

And so Facebook made choices to advance its own success over having a product that was safer overall for people.

Q And the "Criteria to Recommend" column -- so, for example, for columns 6, 7, and 8, it says "Met -- Recommend to Roll Out Immediately."

A Uh-huh.

Q The criteria, are those housed in a document somewhere, or are those in someone else's head?

A So there is a library -- like, there's specifically a tool that lists all the EPICs. So each one of these things has an ID number. We used to call them EPICs, E-P-I-C. There is, like, a tool that houses these, and it includes the criteria for, like, what allows you to turn it on.

And the idea behind those tools, like, the idea of having, like, specific interventions with criteria, is it meant that you could develop an intervention proactively and, like, see what the effects are, like, run experiments to understand, like, what are the harms, like, what parts of the product would be harmed by having this on. And by "harmed," I mean growing slower.

And so there would've been, like, basically, this database, where you could go and say, like, we've already agreed with Policy in advance what the criteria are for turning this on. So that's how it would've worked.

Q So, I guess, who needs to decide on the criteria for turning it on?

A So likely the people who would've initially defined the criteria are the people who developed that intervention. But it is likely that they proposed criteria for turning

it on and then those criteria may have either been accepted by, like, the Policy org or, like, the various other stakeholders, or they might have been suggested to be modified. I don't know on an intervention-by-intervention basis.

Q Yeah, I meant generally.

A Yeah.

Q And do you want to keep scrolling --

A Yes, I'd love to.

Q -- through and see if there's anything you think is important for us to --

A This is such an important document. I can't emphasize this enough. Because it actually shines some light onto how complicated -- like, there should have been a team in the advance to the Jan 6th inauguration that was actively monitoring this situation and probably should've turned on a lot of these things weeks earlier.

And the fact that, at 5:00 p.m., on January 6th -- like, I think the "N/A" ones here means that they're not in the process of being turned on. Like, these don't say "on," right? They say "approved for launch." And I'm pretty sure what that means is they didn't have staffing or they hadn't mobilized staffing to turn them on yet. Right?

Q I mean, "break the glass" is something -- it brings to mind, to me, something you want to do quickly.

A Exactly.

And I want to give you an example. Like, if you scroll down to line 13, like, this is an example of, like -- it just frustrates me so much. So, if you look at line 13, it says, "Not met. We don't currently" --

Q This is column E.

A Oh, actually, never mind. I misread this slightly. So -- oh, never mind. I read that wrong. I thought it was a thing where they said they didn't have enough

staffing so they couldn't do it. But it says that they haven't reached that point. So never mind.

But the point, though, is, like, that "N/A" column, so the one where it says, like -- I think that was originally "Status." If there had been more engineers staffed on this, like, if someone had been paying attention in December and been like, hey, this could be a crisis, they would've pulled engineers off of other projects, at least temporarily, to make sure that these things got turned on. Right? And that's what's so frustrating about it.

I want to call out one more thing. So go down to row 11, so that's RE57. So it says, "Filter delegitimizing entities from recommendations."

So something that may not be obvious is, Facebook has a huge amount of influence over what communities grow or die on the platform. And a thing that was a small scandal last year -- so this was in 2020 -- in Germany, it was found that 65 percent of people who joined neo-Nazi groups in Germany joined them because Facebook recommended them to them.

And so RE57, that's row 11, "Filter delegitimizing entities from recommendations," it's so fascinating to me that this has not been turned on, because -- like it says, "Investigating." They don't know what the delegitimization terms are. Right? That's the kind of thing where --

Q I'm sorry. Can you walk through that a little more for us quickly?

A Sure. Totally.

So let's imagine you knew that there was a large movement, like the Stop the Steal movement, happening and that they had a bunch of terms, like, they had their own code, right, like, their own, like, words that had a meaning to them specifically. Facebook knew that there was a problem of people changing the names of their groups to those code terms. And you could imagine a world where if a group changed its, like,

name to something that is evocative to that movement, that would be a tip-off that maybe that group was contributing to the problem. Maybe Facebook should not be pushing more users into that group. But --

Q So, for example, Facebook took down the Boogaloo Boys.

A Yeah. Okay.

Q And they claimed that they were very responsive to attempts to evade those bans. And is that the kind of thing that would've required someone to look for, you know, the kind of terms that you're discussing that are well-known in the movement?

A And it's one of those things where -- I want to be really clear with people, because I think this is not necessarily obvious if you don't work on language-related things.

If you have a name like "Boogaloo Boys" and you change one of the O's in it to a zero, you will no longer match the systems for removing those terms.

Q And is this related to regular expressions?

A That would be related to a regular expression, yeah.

Q Okay. So maybe that's mentioned in a different column. Can you go into a little bit about that?

A Sure. And I want to actually specifically call that out. I'm guessing the regular expression they were using was not dealing -- so there's this concept known as leetspeak. So leetspeak is where you replace letters with numbers. So, for example --

Q "Leet" could also be, like, upside-down 7, numeral 3, numeral 3, T?

A Yeah. That's an example, yeah.

And so, when you do a Reg X, these Reg X's are really referring -- I'm guessing, are referring to -- let's imagine you had a name that was, like, "Team BoogalooBoys," and you had one that was, like, "BoogalooBoys Squad," and you had one that was "We are the

BoogalooBoys." All those things are different text before and after but the word "BoogalooBoys" remains a single word.

So the Reg X's were almost certainly written to catch permutations like those. They were almost certainly -- I have trouble believing they would ever be written so rigorously that you would also catch people going in there and replacing, like, one of the O's in "Boogaloo Boys" or two of the O's in "Boogaloo Boys," right?

Q Sure. So --

A Yeah, go ahead.

Q -- going back to row 11 --

A Uh-huh.

Q -- this says this is an option that Facebook did not have turned on --

A It didn't turn it on.

Q -- as of 5:00 p.m. And the EPIC is to "filter delegitimizing entities from recommendations."

A Yes. Yeah. And so that means Facebook was still recommending those groups as of 5:00 p.m. on the day of January 6th.

Q Even if they took down the big "Stop the Steal" group --

A Yes.

Q -- early on, there were other groups that may have had these names in them, and Facebook had the option --

A Yes.

Q -- to turn on this EPIC and not recommend them --

A Yes.

Q -- but it had not done that as of 5:00 p.m. on --

A That's what this document implies. Yes.

Q Okay.

Is there anything else you want to talk about here?

A Let's see if there are any other ones that are interesting. Can you scroll down further, please?

Okay. I think that's it for that first table.

Q So, before we get -- oh, I'm sorry.

A Oh. Give me a second. I'm trying to understand -- oh, that's interesting.

Can we look at CP2, please? So it says, "New Opportunities." So what this one is saying -- so it says: Instagram, launch violence and incitement filter for hashtag top at probability 25.

So, like we said before, you know, there's this thing of, it's very hard, especially for super-short things -- so, like, a hashtag is very short -- to be able to guess accurately, is this a good or bad hashtag? Right?

So Facebook had this intervention, which, for the explore tab, reels, and the recommendations, it was on since November, that if you had a 25-percent probability of being a problematic hashtag, that media did not show up in those spaces. This is saying: Hey, we should be also doing this for the top hashtags. I'm guessing this is a part of Instagram.

And it looks like it was met, but it says, "Status: N/A," so I'm guessing it wasn't staffed. So that's an example of something where you could've had it on since November, because you left it on since November for the other ones.

Let me see if any other ones are interesting.

Oh, interesting. Give me a second.

Okay. I think that's all I want to comment on.

Q Okay.

Now let's go to document 5, which I believe we -- if we haven't designated it as exhibit C yet, let's do that.

So now we're at document 5.

A Uh-huh.

Q You mentioned this is a live -- that that document -- or that exhibit A was a live document.

A Uh-huh.

Q Is this a later -- what is exhibit C?

A Interesting. I would love to see the original name of this document, because we can go --

Q It's called --I'll read it for you. It's called "Capitol Protests -- BTG."

A Okay. Let's see. I'm pulling it up on my computers real quick.

It's very possible that I photographed that document at multiple points. Yeah, looks like I have two copies of it. Let me see what the timestamp is on the other one.

Ah, interesting.

So I'm guessing this one -- I would have to go back and double-check. So I have one version of this document that is from January 6th, and I have one version of this document that is from January 7th. And so I am guessing that the version you're looking at -- let me double-check it real quick. Let me open the January 7th one.

Q Okay. The colors in columns F and G in exhibit C are different than --

A Yes. Just give me 1 second. It goes a little slow.

Okay. So the one that you're seeing right now that has the green, like, criteria to recommend, the timestamp in the upper-right-hand corner of that in the unsanitized version is noon on Thursday. So I'm guessing that is the next day.

And let me look at --

Q And that is exhibit C. That's the one we're currently looking at.

A The one that we're currently looking at, yeah. And I think that it's a great document from the perspective of, it allows you to see that they were able -- a bunch of these, now, the status is "live," right? And so, by the next day, January 7th, they had actually turned on many of these things.

God, that's so frustrating.

Q Okay. And if you look through any of these --

A Yeah.

Q -- there appear to be more than there were --

A The day before. Yeah.

Q -- the day before.

So are there any that -- take your time. We can scroll from page 1 to 2, or you can have it up on your computer. But are there any that you think you should draw our attention to, knowing what you know now about when -- refreshing your recollection of when you took this photo?

A How interesting. Interesting.

So, on -- I guess this is under "Phase 2, Follow-Up Action," where it says there --

Q So this is page -- I believe this is page 5.

A Yes. So, if you look at row 3, this is RE39, this is a great example of choices that Facebook made, right, where they made it because the other configuration -- like, whatever their baseline configuration was led to more growth.

So it says, "Mix shift to increase average NEQ," which is news quality, "of news in connected feed. Demote low-NEQ news, and boost high-NEQ news."

So they could've done this the entire time, like, basically, coming in and saying, we have a way of approximating low- and high-quality news, we're going to give more reach

to higher-quality news and less reach to lower-quality news. And that's a great example of, like, they knew that that was a thing that they could do, and they chose not to.

Oh, wow. Oof. Next line, line 4 --

Q Line 4 of the same table?

A Same thing, so it's RE1 -- says, "Remove all feed models for civic posts that use commenting or sharing as positive signals."

I think just a meta-point across all these things is, if it's true -- like, you'll have to go confirm this -- that those numbers, like, RE1, RE39 -- those are catalog numbers, right? In some ways, I'm guessing, the smaller the number, kind of, the more original that intervention was.

And the fact that this one was called "RE1" -- and what it's saying is, for civic content and for crime-slash-tragedy content, there are factors in how the content is prioritized in the news feed, where content is promoted based on the probability the viewer will comment on it or reshare it, like, they'll react to it.

And the second part of that thing says, "Back test results suggest this lever may have been more impactful on hate and graphic violence -- example, US-" -- so it's, like -- I forgot what the acronym stands for, but, like they said, it's like user reports -- "than initially expected."

And so, earlier in this interview, I described how they made the change in how terms of content was prioritized, that meaningful social interactions thing, where, because it's easier to elicit hateful reactions or angry reactions from people, that you end up inadvertently giving the most reach to the most extreme content.

And so it says here, "Priority P2: Wary of negative ecosystem impact." What that means is that, you know, in a world -- like, we talked about this before. In some countries, 35 percent of all the things viewed are reshares. So, in a world where content

is not prioritized based on its ability to, like, elicit a reshare, you end up decreasing the amount of content available in the system.

Line 5 is interesting. So a thing that is not obvious, likely, is that -- and this is covered in things like -- I don't know if you guys have received the documents about COVID misinformation, but that research is really, really well-done in terms of explaining how there are differential impacts of these systems.

You know, some parts of the system -- like, a very, very small number of users -- like, in the case -- so, in the case of COVID, if you cap how many comments everyone in the system can do at the level that the P99, so the 99th percentile, does -- so maybe that's, like, 30 comments a day -- you end up substantially decreasing the total amount of COVID misinformation comments, because a very small number of people spread the vast majority of COVID misinformation. Right? Like, maybe the P99 spreads 50 percent of all the COVID misinformation, something like that.

So row 5 says, "Demote users who repeatedly report misinformation." So this is an RO, a repeat offender.

A very small fraction of people are the mega-amplifiers of all these problems. Like, they go and they repost something into 50 groups, that kind of thing. And so, if you come in and say, hey, there's a very small number of these users who have been caught spreading lots of this stuff, let's demote them. And Facebook is very wary to do these kinds of interventions, even though they are very impactful.

Let me just -- give me just a second, because I'm scrolling up and just seeing if these other things --

[1:08 p.m.]

BY [redacted]

Q We're on page 2 or 3.

A Just give me one second. It's page 2. Why don't we go to page 2 real quick.

Q Sure. So I am looking at row 11, column G. It says, live for 2020 reg exes pending from new reg exes?

A Ah, God. So this a great example of how had they staffed more people prior to January 6th, they would have been ready to go on January 6th. So remember how in the previous document we saw that they hadn't turned this on because they didn't know what terms to use? What this, it looks like, says is that now -- like they basically said let's revert to what we knew the bad terms were for the 2020 election. They still haven't done the research. They still haven't staffed enough people to figure out what the new code terms are for these groups. Like, it's just -- it's shocking, which means that Facebook was actively recommending groups with, like, the -- with -- that where promoting these problematic things.

Q Do you want to mention anything else?

A I do. I do. I just want to -- give me one second. I am reading some of the ones below it. Okay. Let's talk about RE 53. That's actually really important also.

Q On page 2?

A Oh, no. Oh, sorry, not that one. I want the one -- oh, actually, let's talk about that real quick. That one is actually even more interesting than I thought it was. So before we were just talking about in terms of being not prepared, I want to explain why these specific ones are problematic. So if you look at column D for row 11 -- so,

remember, this intervention is called preventing pages from changing their names to de-legitimatizing terms. A strategy that is used by information operations is that they build up large pages. The MIT Tech Review did a really good article on this in September of 2021. They will build up these large audiences pretending to be specific target groups. So what that article details is things like a huge fraction of all the Evangelical Christian groups are not run by Evangelical Christians. They're run by foreign actors. A huge fraction of African-American groups are not run by African Americans, they're run by foreign actors.

And the thing that they do is they build up these large groups using something called virality hacking and growth hacking, where they'll go in there and they'll say find content that went viral in other places, and they'll start reposting it to tier groups. That's only the most click-baity, most-likely-to-spread-on-Facebook stuff.

And so, they build up these very large audiences, like maybe half a million or a million people, and then when they want to use them for their original purpose, they change the name of that group and change what the content is going to that group, because they still have the distribution channel.

And, so, what this is saying is we know there are lots of these groups that might be -- might have seem innocuous before. It might have been about cute pictures of dogs for all we know. And that the strategy the information operations uses, they go and take that group that they built for a different purpose, and then they rename it and make it now into a megaphone that can blast out this problematic content.

And that's why this civic intervention matters so much, because it is a known technique used for these situations. Fit 12, which is RE 57. This is another one of those things around Facebook actively promoting de-legitimizing entities. And this is, again, a problem that they turned it on for the terms that were relevant in

November 2020. They had not updated it, it looks like this, based on this with the new bad terms.

Q So they had been paying attention to the election --

A Yeah.

Q -- but they turned their attention away afterward?

A Yes, exactly. And it's actually slightly worse than that because it means like they were still promoting entities that were causing these problems as of the day after January 6th. Right. Because they still hadn't gotten the new addresses. And the line below that is amazing in terms of -- so line 13 is RE 8. So this is, like -- so it says, "key word-based emotion of probable violating PSR at 50 percent strength." And the -- so it says in the description, "demotes storm the Capitol PSR at 50 percent strength with 70 percent" --

Q PSR meaning?

A I think that is -- I am looking at something about probable something review, but I don't know for sure. We could check across the cache. But the reason why I was commenting on it and what was amazing about it was it's still not done. It says execution underway on the day after the election. Give me 2 seconds, I'm just looking at the rest of it. God, can we talk about line 9? So it says PE 16, stop boosting content from nonrecommendable groups in feed.

So, you know, I don't think that was on the chart from a day before. I don't remember seeing that. So what this line means is that Facebook knows there is some groups that they think are bad enough that they are not recommendable, right? And the -- and yet, that content still gets boosted by other models in the news feed. And so the description says, removed several boosting models for nonrecommendable groups in news feed. So given that was not on the chart yesterday, I don't remember seeing it for

the Jan -- the day of Jan 6th chart. Like that means they had groups they knew were bad enough they weren't going to recommend people join them, but they still were boosting the content from those groups.

Q So these are groups that Facebook has decided, for whatever reason, we're not going to recommend people join these groups?

A Yeah.

Q But groups are places where people post a lot of content?

A Yep.

Q And that content can be shared either to someone's friends who opted in that group, or even to someone who's friends with someone who is in that group even if they didn't post it or otherwise?

A There is a couple of different things. One is Facebook sometimes shows you content from groups that you did not join because a friend of yours interacted with that content, right?

Q Okay.

A What could even be that, like you are not even a member of this group, which Facebook knows is not good, and that you might still be sent content from that group because your friend interacted with it. So, yeah, like the -- Facebook had boosts that were giving more description of this content, even though they knew that the groups themselves were bad.

I would like to go down to line 15 in that same chart. I think it's page 3. This is another really great example of something that I don't remember seeing on the chart the day before. So it says, demote deep reshares, two plus of content classified civic, including crime and tragedy. And the description is "demoted all civic reshares except those that were reshared directly from original post."

So the -- I mentioned this earlier, this idea of you are not choosing good and bad content, right? You are just saying, Hey, you know, if you get beyond friends of friends, you should, like, let's -- like, that we shouldn't let things go, given that we know that a nonrepresentative set of content gets reshared more deeply. Like, you know, in this case, they're not cutting the chain at two, they're just demoting it after two. But this is also a known thing that substantially decreases misinformation, it substantially decreases violence. I think they knew about that for a full year before this happened. Like we have documents on that one. And the fact that it wasn't live before Jan 6th is, I think, negligent.

Let's me see what the other ones are. Oh, wow, if you go on to line 16, it says PE 25. So it says LVV boost cap.

So we discussed yesterday the idea like one of the suggested things that could be done was moving from boosting live videos at like 850, moving this down to 60. Or excuse me, yesterday, it was from 850 down to 833 -- or excuse me, 83. And, today, they have now changed that to 60. So it's even lower, because it was contributing to ultrarapid virality for several low-quality viral videos. The same thing.

It looks like now that's live. Give me 2 seconds. Oh, interesting. Row 18, P22, another one that I am amazed it wasn't on before. So PYML is pages you might like. And, so, that's the recommendation system that promotes pages to people. And so this one says is prevent low-quality and misinformation pages from growing the audience and becoming viral.

So what this is saying was on -- I don't remember seeing this on the day before. So that means on January 6th Facebook was actively promoting pages that they knew were misinformation pages, or that they were just low-quality pages. But by January 7th they had turned off that.

What are other ones that are good? This is another great one. Go to row 20. It says PE 27. It says "lower threshold for hateful comments by applying tier 1 at-risk countries methodology." That's ARC. So tier 1 at-risk countries, technically, Facebook was a tier 1 at-risk country because it shows --

Q Facebook was or the United States was?

A The United States was. Excuse me. Technically, the United States was. Like, they look at the ARC. It's a great example of how sometimes countries are put in tier 1 for instability reasons, and sometimes they are put in there for political reasons.

The United States was in the tier 1 adverse country classification at least in 2020. It might have been that by January 6th, they were no longer tier 1. That's totally possible. But the methodology for in at-risk countries, they have a slightly slower threshold for when they qualify a post to be demoted for being like a hateful comment. And, basically, they had this ability, they could have done this on January 6th. I don't remember seeing it on the January 6th list. But this is just saying we already know that this is the threshold we use in places that are unstable. Let's turn it on for the United States as well.

Line 21, this is another example of like something I didn't see on January 6th. I think this document is really, really problematic, because this document shows how far behind they were on January 6th. Because even a day later, this is a much more rigorous response than it was the day before.

So the thing I wanted to flag on this one is if you go to row 21, RE 41, it says, "designate U.S. as a temporary high-risk location." So there are -- you know, like I said before, Facebook doesn't, you know, slowly get more rigorous. It likes to step-wise -- it acts in a step-wise manner. You know, it doesn't -- it knows all these ways to make the system safer, but it doesn't do them until like literally chaos is ensuing, right? And one

of those things is that --

Q You mean, this is a regular occurrence on Facebook?

A So as we saw in the previous ones, so it's row 20, PE 27, where it says, "tier 1, at-risk country methodology." There is a bunch of places in the world that Facebook considers to be a high enough risk that they get these special treatments, right? And because Facebook was not considering -- I don't remember seeing this line on the day before, right? So it says "designate U.S. as a temporary high-risk location." The fact that no one inside the company had designated the United States as a temporary high-risk location on January 5th is crazy to me. And so it looks like by January 7th, they had designated the United States as a temporary high-risk location. And because now it had crossed that step, that step-wise, like, criteria, now they are able to act in a more assertive way.

Oh, interesting. If you go to line 23. Oh, actually, let's look at 23. So it says, "turn off profile frame creation." It says "disables new frame creation in the United States. Use of frames created outside the United States in within the United States." I am guessing what that means is that they detected that people outside the United States were creating inflammatory frames. So frames are, basically, like images that you can apply over your profile picture that then get distributed across Facebook. And I am not aware of January 6th-related profile frames. But the fact that it's in here, and that they specifically call out not allowing ones created outside the United States in the United States means there were information operations feeding inflammatory frames that were contributing to the conflict.

And given this is then today's, like the January 7th post, but not in January 6th, means that they were not aware this was happening, or they hadn't acted on it until the day after January 6th. Feature block the admin's -- oh, that's really interesting. Can

we look at row 24, PE 31.

Q Uh-huh.

A So a complex object, I believe, means a group. But I know that we have documents in the cache that specifically describe what is a complex object. So admins, moderators, and top-contributing members are people who either run, like own the group, administer the -- are like moderators in the group, or are the most frequently posting people into these Facebook groups. And a lot of these were disabled, right? So like there is documents in the cache describing what happened to the Stop the Steal movement in December, right?

And what this is saying is that if we know that you were an admin, or, like, you were a moderator, so you were very involved in a group, or you were like a super poster into that group, if your group got disabled for being a dangerous organization -- so I'm guessing, you know, this could be terrorism, this could be, you know, violence-inciting groups. It's saying, Hey, we know you were really, really involved in a dangerous thing, we're not going to let you post more on Facebook for 30 days.

I don't remember this being on the list on January 6th. The fact that they turned it on January 7th meant they knew there was a small number of people fanning this conflict, and that the thing they can do to make the situation safer was to say, Hey, guess what, you can read Facebook for the next 30 days, but you can't contribute to it. The fact that that was not even on the list the day before, I think, is really problematic.

So I'm just going down -- I haven't read this document closely before. It's an amazing document. Oh, interesting. Let's move down to row 26. So this is on the next page, I think, it's RE 46. So it says, "set up a special re-review queue for high-risk contents from our user reports with risk and response."

So one of the challenging things for Facebook is that, like, certain kinds of

misinformation are not considered misinformation until a third-party fact-checker says this is false. So you can imagine a situation where a claim is being frequently reported, like, so that's a spike in violations where you might want to say, Hey, let's look at these -- this concept, or like this thing is getting reported over and over and over again from being problematic. Yes, our first-tier thing said, Oh, this doesn't qualify. Well, we're going to set up a thing within risk and response to look for those things that are happening over and over again, and potentially say, Oh, no, it's actually problematic and take it down.

And in this case, I don't know what the criteria means. It says not met. There were different triggers that made this more relevant for 2020. And I don't know what that meant. It might have been a thing where they did it more because it was around voter disenfranchisement or something, but I can't imagine that it wasn't still relevant in this context. I'm just going down a little bit more. Just give me one second.

If you can go down to -- it says "new levers not previously launched." So row 4, CP 6. It says, "apply the DOI roadblock, read-only access. It says, dangerous orgs and individuals. Apply the roadblock read-only access to approximately 7,000 users. I'm guessing that that was another thing where they came through and found that those 7,000 users were particularly core to fanning this conflict, and they moved them into read only. They could have done that, unquestionably, the day before or a week before if they had had enough people monitoring the situation or they had been a high enough priority.

Q Some of those users there isn't any normal --

A Isn't a what?

Q -- group of users? They had a reason for tagging this number of users in these 7,000?

A So on the description, it says T82639518. That means it is a ticket, so the ticket is somewhere in the system, where in -- I think we should make sure you guys get the unredacted copy of this specific document as well because I know we got you the unredacted copy for Jan 6th. I think you should go follow up with the people on the owner column on all of these things, because you could go talk to those people individually.

So the fact that there is a ticket there means that there was an investigation, a specific investigation that found that those 7,000 people were, for some reason, instrumental in this conflict, and that those people were moved into read-only access. If you followed up with that owner, because there is a possibility that this ticket has already been -- oh, it probably hasn't been auto-deleted unless Facebook deleted it. So you have about an 18-month, I think, time horizon when Facebook automatically deletes thing unless they are, specifically, tagged to not be deleted, which I think is negligent. They should not be allowed to delete tickets, period. Because they're not -- it's just a different level of sensitivity, and they -- I think they use it to cover their misdeeds, if you want a recommendation for the future.

There is definitely something about those 7,000 people where someone identified that they were increasing this conflict. Let's see what the other ones are. Interesting. Okay. Almost done. Just scrolling down. I am now on phase 2, follow-up action. If you look at RE 35, this is row 2, it says "demote unconnected risky entities and their content." Initially, because of the filter pushed back to -- pushed as demote. So this says, "demotes newly created entities, entities with recent name changes or long dormant periods in order to prevent growth hacking via trending terms in the lead-up to elections, also demotes content and events created by the entities."

So this is a really interesting one, because it -- so what we talked about before

people sometimes build up these groups and get very large audiences, and they might not even continue to use them. So I want to -- something that I think is -- I can't imagine anyone outside the company knowing this.

Up until some point in 2020, you didn't have to have an owner of a group. Like you could have had a group that was quite large, and whoever, like, was running that group or running that page, probably not the page, but running that group, they could have left the platform, they could have been removed from the platform, like, maybe they were naughty and got caught and were moved for a different thing. And because the group was not required to have an admin, a random user could claim the group and become the admin, and become the owner.

And so a thing that happened with nonzero frequency was there were people like information operations that would go and find these groups and clean them, and they might leave them dormant, like they might not even continue to post into them, but now they have a distribution channel where they can go and push out content to this large group or this page.

And what this is saying was that I don't even know if this is rolled out yet because, you know, there is no status thing on it. It says, you know, Facebook knows these are risky entities, and either because they are newly created -- so what's talked about in other cache documents is this idea that you can make a group, and every day each individual in that group can invite, I think, it's 2,200 people to the group, right? And what that means is that you can have -- grow groups very rapidly, and then have a mechanism for pushing content out to them. Because I think for the first 30 days, Facebook will deliver content from that group to your news feed, and if you engage with any of it, it's considered a follow-up.

And so what this line is saying is that, you know, if we know that this is a risky

entity, we should demote the ad content in this feed. And the fact that that was not also not live on, say, January 1, I think, is really amazing. It looks like they hadn't launched yet. So remember we talked about the idea of the mixed shift of low-quality and high-quality news? So row 3 in that chart, RE 39, it looks like there is a note on it saying they still haven't decided whether or not to launch it, which is amazing. So they're still over-promoting low quality news effectively because of this.

Oh, wow. They still haven't done -- so remember we talked about before RE 1, that's row 4? They are still promoting content that is based on the probability of someone getting like a reshare or a comment, which they know is a thing that causes a more hate- and graphic-violence content to be distributed and still says it is a low priority because they're worried of, quote, "negative ecosystem impact."

Q What does that mean, negative ecosystems?

A Oh, it means like hating MSI, right? Like literally that will hit MSI, because both the comments and the reshares are considered MSI. Also, in the case of reshares, like, it decreases the total volume of content in the system. RE -- line 5, RE 24, it looks like it's still not on. So if you are a user who repeatedly posts misinformation, you are still not getting demoted as of January 7th. That's amazing.

Q And if you look at column E of this table?

A Yes.

Q It says "priority," and there is P1, P2, P3?

A Yeah.

Q Can you just describe what that means in Facebook's context?

A P2 and P3 like -- P2 and P3 mean that you're basically not doing it, right? Like --

Q Are there also like P0s?

A So a P0, classically in software engineering means you cannot go home, right? Until it is fixed, you cannot stop. And so, you can have a P0, but they are rarely used because they are extremely disruptive to people. P1 means it is likely being worked on. And P2 and P3 effectively mean you're not doing it.

Q Because there is a lot of -- it has to do with Facebook being extremely thinly staffed?

A Facebook is extremely thinly staffed, yeah. Or in this case, they're not doing it probably -- like based on the comments in this doc, they're not doing it because they're not willing to like, say, to take a hit to core MSI, right?

Oh, interesting. That's amazing. If we can scroll over further down. So it says U.S. 2020 Capitol protests, break the glass, violence and incitement, V&I. And it says Instagram V&I break glass.

Q Is this a new document while it's part of the exhibit?

A I think it's a new document. I think when I separated it in folders, I sometimes didn't do it cleanly enough.

Q I just want to clarify for the record. So what were you going to say about it?

A So what I was going to say was let's at -- so it says, at the beginning it has some bullet points. It has, like, one, what is your product? Two, what are we seeing that is triggering this proposal? And it says V&I, FRX report. So V&I is violence and incitement, and FRX is the system where, like, people report things. It's like someone goes and reports visa content, that's an FRX report.

So they are saying a couple different things. One, the reports are spiking in terms of, like, people reporting violence and inciting content. And then when they did a manual review of the top-reported posts, they saw lots of violence. Oh, wow. I don't

know what CAD, C-A-D is, but it says "CAD pipelines." I am thinking it's civic something. It says pipelines are behind as of 9:00 a.m. and -- I am trying to see if there is a timestamp on this one or any of them. Two seconds. I was too clean in photographing these. It's too bad. Because there's not a time -- I don't have a timestamp on this. But yeah. Just give me one second.

Q So this is also a working document, it sounds like, you think?

A I can't imagine it not being a working document. Actually, no, I know I did do a timestamp on that because I have in the folder. Okay. The timestamp on all of those in this document are from the 7th of January, according to like the excess data on the photo. So the thing it says is "CAD pipelines are behind. As of 9:00 a.m., we do not yet see a spike in high-precision classified media and many false positives." So that means is that the dynamic systems are not working fast enough to remove this content. It says --

Q Were those automated systems?

A It might be automated, or they might be like taking in new signals. So the sub book plan on that one says, "violence and inciting media classifier on IG is new." This is a recurrent problem, just so you guys are aware of this. Instagram regularly runs behind of Facebook on these safety systems. Part of why I am, like, hesitant to break up Instagram and Facebook is just that like there are so few industry professionals working on these things that like, you know, even keeping Facebook safe versus keeping Instagram face, regularly Instagram didn't have sufficiently advanced safety systems. It says, "will take some time to see if the classifiers actually capturing reported civic content." So basically what this says is, because they underinvested in violence and incitement content destruction on Instagram, they don't even know if the violence and incitement system on Instagram is working or not, which is crazy.

If we go to bullet point three, it says "why business as usual is failing question mark." And it says, hashtag top, which means like the top hashtags, feed-in stories. So feed is one and stories is one, surface area, or content can be shown. It says they have no existing violence and incitement protections. So if you see violence and incitement content going through those places, there is nothing reducing that content. So it says what are we proposing? Recommend filter violence and incitement hashtags from the hashtag top at probably 25 in the U.S. So we discussed this one before. And what it is saying here is this intervention has already been alive since November, and explore reels and recommended content.

I am guessing that that was used. So if those hashtags were trending, it meant that you had a distribution menu space where you were not removing violence and incitement content, and you were giving much more reach to it if it was one of those trending terms, which on January 6th, a lot of it was around the Capitol riots. Let's see what other interesting things here. Reduce comment demotion thresholds.

Q You are looking at the table at the bottom?

A Just give me a second. This one has like a -- it's harder. Oh, interesting. So one of the things I find interesting about this is -- so did you see the column "external defensibility"? So there was a number of documents in the cache that relate to defensibility. Like defensibility is such an interesting concept, because -- so Facebook is so scared of being blamed for taking out content that they shouldn't have taken down, or demoting content that they shouldn't have demoted, that they allow tons and tons and tons of bad stuff to get through. And things that were known to be good at intervention -- so I will give you an example. They know that a very small number of people are hyper sharers, right? So a very small number of people do a huge fraction of all the reshares. And there is something called "sparing sharing," which is a great set of

docs, if you could look it up. And sparing sharing just said, Hey, we know that people who are hyper sharers -- these are people who share, you know, like 20 things every day or more -- they pose a disproportionate share of all the misinformation on the platform. If we just slightly demote those people, we end up having a huge reduction in misinformation.

And they said given that there was no proof that people who were hyper sharers were bad people, you couldn't do that intervention. It is considered not defensible. The fact that on this list, they even have a column for external defensibility shows you how much Facebook was biased towards not acting, because they're even assessing right there the defensibility.

So looking at row 5, it says "downrank all IG lives," so they're outside of top four in-stories tray. We saw that on the other, on Facebook, that live video was playing an outsized role in spreading this violence. What this is saying is given -- I am almost certain they have almost no protections on IG live, because there was a recurrent problem on Facebook all the time. But they're saying given we have so little visibility into this as a problem space, let's demote it enough so that they're at least not in the top four things, they're not immediately visible in your stories tray.

And this is a great example. So under "external defensibility," they're saying, like, Well, we're not going to be able to explain this to authors and viewers. And it's like, well, what if you we explained to them you just can't do your job, and so you're not going to, like, make this as obvious during a time of high crisis. Like, that should be defensible enough. But the fact that they described this as, like, you know, authors and viewers wouldn't even tolerate this for a day or two, it's just -- it's characteristic of Facebook's approach to safety.

Q And that box is red?

A It's red.

Q Yeah. Do the cards on this table mean something different from the cards on the previous tables?

A I don't think so. This is a very frequent communication.

Q The red, yellow, and green?

A So green means good, yellow means so-so, and red means bad.

Q Okay. So this is a --

A Yeah. And so, part of what's frustrating for me about these are ranked as red is if you look at efficacy at mitigating risks for rows 5 and 6, this is around Instagram live content. The first thing they say is live risk not established. And that is a great example of the cost of understaffing things at Facebook. Lots and lots of problems are vaguely known to be problems, but because Facebook is not adequately invested in those -- has not adequately invested in those systems, they don't even know how to assess the risk on things, because they haven't -- they don't -- they don't even know how dangerous things are because they haven't staffed enough people. It's crazy.

Q Is it also possible that people might not study something because they don't expect it will move forward even if they find results?

A I'm sure that's true. I'm sure that's true. There's a thing of, like, you know, if it impacts growth, if it impacts, you know, depth of engagement, you know, if people have a sense that that would happen and they know it would be a huge fight, it makes it harder to actually be able to intervene there.

Hmm. Interesting. I am just looking at the rest of the document. I don't think I have any other specific comments on this. I am happy to answer questions, but I don't else interestingly.

Q So on the next page, if you look at these graphs here. It says "violence and

incitement to high classifier score," and then some various graphs. So do you know how they calculate violence and incitement?

A I don't. That's actually one of the things that I think is really problematic about Facebook is that -- so violence and incitement, they had some examples of content probably, and then they trained an AI, an artificial intelligence system to try to find other things that were like those examples. We're never shown the examples that are used for training, and we're never shown like what the performances of these systems. And so, I don't know exactly what they used as examples of violence and incitement. What these graphs show is that in the days running up to January 6th, it looks like there was a spike on Jan 2nd; maybe Jan 2nd or Jan 4th, where a lot more violence and inciting content was distributed. And I don't know what caused that. And that's in reels. So reels is like the TikTok one. There was a spike. And there was a spike -- I guess there was a spike kind of across the board. I am looking at what the different ones are. So it looks like on posts, there was a spike in -- so there was a spike on Jan 5th, both in reshare -- or maybe Jan 4th, in reshares and posts for violence and incitement.

And then in explore, there was -- there was very clearly -- so just for context for people, sometimes when you see these very, very sharp spikes, at least there was one particularly high distribution post. So, for example, like, when you look at explore, those two little spikes under post impressions, that would be like there is likely two posts that got extreme distribution on those dates. What's interesting is there was a different thing that happened on, like, Jan 5th, or late Jan 4th, where it looks like some content got extremely highly reshared. So like there is something about it that the algorithm caused it to boost that content.

Q Is there a team or a person who is supposed to be watching this data?

A So that might be a thing that would get watched. So if -- so remember

what we were talking before like what was the role of Samidh? Like what did Samidh have the ability of doing? So someone would have had to make a call to say we need to staff up -- like what would be more conventionally called a war room. They call them IPOCs, like we discussed before. Someone would have to make that call saying we need to staff up an IPOC. And Samidh would have been that person in the world where civic integrity had not been dissolved.

It is likely no one was, specifically, looking at this data unless it was maybe strategic response, like -- or yes, strategic response, unless there has been a war room staffed up for January 6th. I am assuming there wasn't a war room staffed up because of how big the difference was between the Jan 6th version of this document and the Jan 7th version of this document. I don't know that for sure, but given how much the needle moved in even a day, it makes me think there wasn't someone watching before Jan 6th.

Q I am about to wrap up. There is a lot we're not going to be able to get to today.

[redacted] I think you are really muffled there. So you might take a break for a second and then start over.

[redacted] Okay. Am I better now?

[redacted] Yeah. You're good.

[redacted] What I was saying was I think we're going to wrap up on these three documents, these three exhibits and take a break, because now is approximately the time we're going to take that break. And then we'll come back and we'll talk about, I think, the soft actioning. Mark's feedback on soft actioning proposal. And deck presented to Mark's document. And then depending on how long that takes, we have a lot -- we may leave the rest of the documents to another time and turn to some follow-up

and some wrap-up question?

Ms. Haugen. And I happy to give you guys as much as time as you need. So if we have to do this multiple times, totally, fine. I am here to be helpful.

[redacted] That's great. We appreciate it. I am sure we'll take you up on it. But for now, I think we're going to give everyone a little bit of a break. Drink some water, catch your voice back.

Ms. Liu. Okay. So be back in 30?

[redacted] Is 25 okay?

Ms. Haugen. Yeah. I can do 15. I can do whatever is useful to you guys. Just let me know.

[redacted] Yeah, instead of mentioning an actual timeframe, let's say an actual time.

[redacted] Yeah. That will be great. How about 2:15?

Ms. Haugen. Perfect. Okie dokie. I'll be back.

[Recess.]

[redacted] All right. Back on the record.

BY [redacted]

Q It's 2:19 p.m. It's still November 22nd, 2021, and no one new has joined in, and the same people are named. So, now, let's turn to Document 6, which we're going to designate as Exhibit D, if we could pull that up.

It is called, I believe, the file is called, "Mark feedback on soft action proposal presented to Mark." If you could scroll up to the top. There you go. So --

A We should actually give you the unblinded version of this document because the person who wrote it is a director, and our general rule has been, like, if you are a director or higher, your name is not protected.

Q Okay. Do you remember who wrote it?

A Yeah. I can get that for you right now. Give me a second. Mark --

Ms. Liu. Hey, Frances, can you also send me the location so I can have the team create a non-redacted version?

Ms. Haugen. Can you take a note on that, Libby.

Ms. Liu. Sure. That's fine. I'll do that.

Ms. Haugen. So the person who wrote that is -- one second. Anna, A-n-n-a. Stpanov, S-t-p-a-n-o-v.

BY [redacted]

Q S-T what?

A S as in Sam. T as in -- I don't know, Toronto. E as in echo. P as in papa. A as in alpha. N as in November. O as in October. V as in Victor.

Q And did you provide this interview into the committee?

A Yes.

Q And so, it looks like this document is a number of different things put into one. So the few two pages are notes, and what -- how are those two pages relate to the third page and what appears to be a presentation coming from page 3 through 35? If you need, you can scroll down a little bit.

A I am going to -- I have the original, and I am opening just the full thing. Just give me a few seconds.

Q Sure.

A All right. So the first couple of pages says there is like -- there is like basically a note written, and then there is like what's it called, a screen shot at the bottom of like the second page of the title slide, which says "soft actions task force. Big ideas to reduce prevalence of bad content in news feed." And it says "Mark April 9th," which

means that it was presented to Mark on April 9th. The note overall is from April 22nd. I guess there was a lag in, like, maybe following up on the feedback, or on just writing up the notes and just sending them out.

The remainder of that document is the slide deck itself, and it is -- the title of it, as you can see, is "soft actions task force MZ," which is Mark Zuckerberg, April 9th final, which means how it was presented to Mark. Did I answer your question?

Q Yeah. And so was this -- did you refer to this document earlier, or were you referring to a different document?

A Yeah, I was referring to this document earlier.

Q And are these changes that would have affected the Facebook in the United States?

A It is ambiguous. So there are other documents that list the United States as an at-risk country in 2020. And given that this doc was specifically done in April -- so if you look at the first page, it says, our top-level goal is to take a set of bold integrity product ideas to Mark by the end of March, and then to launch by the end of H1 2020. So that was meant to happen in place, ready to go for the U.S. 2020 election.

So I am assuming these would have applied to the United States. Because the United States was classified as an adverse country in tier 1, and it was meant to be to shipping those available as in the -- so, yes, a bunch of these would have been applied in the United States in 2020, if they had been done.

Q And how do ideas get generated for this slide?

A Oh, great question. So I participated in that process directly. So they did a pretty broad call for suggestions in, I believe, January of 2020. So Anna Stpanov, as I said, is a director in integrity, and she had done a big project around defensibility in Q4 2019.

So that's basically going through all the integrity systems and understanding, like, how they worked and assessing them for quote defensibility. And her next big, like, cross-org task force thing like this one, which was the soft actions one. And they -- basically anyone who wanted to pitch an idea to come to this, like, pitch fest, where like everyone got like 5 minutes, I think, or 10 minutes to pitch ideas. And I think over the course of the entire process there was probably on the order of 60 people involved.

And so of those ideas, if you scroll down to, I think, it's like slide 4 or 5, it says purpose of our discussion is the title on the slide. And they have, like, some buckets like for this idea of like a lot of these are around expanding.

Q I am sorry. Before we get too far. Can you explain what "soft action" is?

A Oh, sure, yes. Let's start there. So a hard action is something like removing a person from a platform or removing a piece of content from a platform. A soft action is around, say, changing the dynamics, how the system as a whole functions such that the system itself is safer in general.

I am a big proponent of integrity, like changes to Facebook that are in the soft actions direction, where you say things like, remember when they're going through the line by line on -- okay. For the Jan 6th break the glass measures. Like one of them was, like, demoting stuff that is reshares that are more than two hops down. Right. So that's an example of a soft intervention. Like no one is, specifically, removed from a platform. It's just the system as a whole is made more effective. Some of these are not on the side of content distribution, they're around the idea of adding friction to the platform. So informative friction is like you educate people before they share, and that reduces the spread of misinformation or other viral extreme content.

Q And soft actions are one of the things that people were considering that

would help reduce misinformation or radicalization on the platform?

A For violence and incitement too, or graphic violence. It's basically anything where -- of the harm types that Facebook knew were disproportionately spread through the growth virality, like reactivity, optimized parts of Facebook, a soft intervention comes in and says maybe there's like a safer configuration of the product. Like instead of focusing on this is a bad piece of content, this is a good piece of content, it's, like, could the product as a whole be safer?

Q Okay. So you were speaking about the purpose of the suggestion before I asked you something.

A And so it looks like of all the suggestions that were produced, and there was a large number of suggestions, it looks like they kind of broke them down into larger themes, things around probable violating, which means like we talked before about the idea that Facebook sets the bar so high for hard interventions on content, you could impact vastly more content by looking at probable violating instead of violating. Informative friction. That's proposal 7, which is around, basically, like, you know, telling people, like, Hey, like this content is known, is like misinformation, or this content comes from maybe a page that has spread a lot of misinformation. So you can imagine various ways you could educate someone to, like, decrease the chance they would then share it. That adds friction to the process. And then they have more, like, they say aggressive concepts. And so we have to go through them one by one to see like what those individual interventions are. Those are kind of like thematic groups.

Q Let's take a step back and talk about the note which --

A Yeah.

Q -- which is on pages 1 and 2.

A Yep.

Q From what I think you said, this is the report back from the presentation?

A Yes. Yeah.

Q Is there anything that you think is interesting about this report back?

A So, if you look at very first slide, so it has -- so there is kind of a heading that says, "attached to this material we showed with Mark" -- this is what we presented to Mark Zuckerberg -- and below is the summary of the guidance we received. So they talk about, like, what to do, the -- oh, God. Amazing. So for the first one it says, for probable violating. He talks about the idea of, you know, you could have it as a break glass measure you could use in an extreme situation, but you can't use it in general. And the secondary thing is, if you'll see the second sentence says "not supportive of account level demotion." It said we will not be launching this. So remember one of the break glass measures was things, like, let's demote content from people who are serial repeat offenders. Like people who have over and over again spread misinfo. When you say not supportive of account level demotion, that's saying you're not going to do that kind of thing.

So the second one is downstream model deprecation. This is the one where I said that I thought this was really, really inappropriate, Mark's decision here. So it says, "Mark doesn't think we could go broad, but he is open to testing, especially in at-risk countries." We wouldn't launch if there was a material tradeoff with MSI impact. And so it says here for a bit more context, this is an extension of launches done earlier this year on health and civic content as well as some launches going out this week to specific to at-risk countries, like Ethiopia and Myanmar.

So what this means here earlier in 2020 for, like, when the riots started happening around Black Lives Matter in 2020, they came through, and for content that was classified as civic -- and remember there were other documents in the cache that talked about the

idea that the classifiers missed the vast majority of content. For content that was classified as civic, they turned off the downstream factors for it and for health-related content, which relates to COVID content.

What this is suggesting here is that because the classifiers are so ineffective, like they're not written in most languages. Even in the case of Arabic, like not all of the dialects or variants of Arabic are similar enough that the Arabic classifier works. Or in the United States like because these classifiers missed so much content, it's saying in situations where you are at risk of violence, let's turn off the downstream models, because you know that when we turn them off, we get less violent content, less misinformation, less bad content in general, which was why it was turned off for civic content and color content in 2020.

But this is saying, Hey, we need more coverage. We can't hit, you know, 5 percent of civic content, which is probably what the classifiers were hatching. If we turn it off in general, the system as a whole would be substantially safer. But the problem is that directly trades off with this meaningless metric, MSI.

And I am guessing the reason why they didn't do it is because lots of lots of people's bonuses are tied to MSI. And it really, really frustrates me because I think substantially less sensationalistic content, significantly less violent content would have been distributed on Facebook had this been actually put in place. And it wasn't put in place because they wanted to protect a meaningless metric, MSI.

Q And how does that relate to January 6th?

[2:33 p.m.]

Ms. Haugen. January 6th was made worse by having hyper-distribution of extreme content. And much less extreme content would've been distributed. The movement -- so we talked about before the idea that if you are invited to a group but haven't affirmatively joined it and haven't affirmatively rejected it yet, you will get delivered content from that group for 30 days, and if you engage with any of it, it will be considered a follow, right?

Having this in place, the downstream models -- like, there were many other documents in the cache that describe the impact of the downstream models for MSI. Having those in place, combined with some of those policies, meant that people who never intentionally joined, say, the "Stop the Steal" movement got invited to it, and then these models selectively gave more distribution to the most extreme content.

So I think, if this had been in place, there would likely have been meaningfully less escalation before the 2020 election and meaningfully less escalation in the period after the election in the lead-up to January 6th.

BY [redacted]

Q And this was teed up to Mark Zuckerberg?

A Yes, in April of 2020.

Q And do you know who decided it was important enough for him to weigh in on it?

A I'm guessing that -- I can't imagine Guy Rosen did not get to see this proposal before it went to Mark. And I'm guessing that Anna --

Q I'm sorry. Wait. Wait, wait. Give me 1 second, Frances.

A Sure.

Q I think you were breaking up -- for me, at least. I don't know if for everyone else. So could you just start over again?

A I can't imagine that Guy Rosen didn't see this before it went to Mark, right? And I'm guessing that, given that these are changes that have material tradeoffs, right -- like, let's say you downstream MSI, like, you turned off downstream MSI in at-risk countries, that would harm, for example, John Hegeman's team. I'm sure there were lots of people on that team whose bonuses would've been affected. Right?

And so anything where you have a tradeoff across orgs like that is going to -- even if you didn't take it to Mark initially, it would likely get escalated to Mark, right? Because it's a conflict between --

[Audio interruption.]

A Sorry. Can you say again?

Q No, no. Finish your thought.

A Because it's a conflict between, like, safety and growth, right?

Q You've mentioned bonuses a couple of times.

A Uh-huh.

Q So are you saying that MSI was a useful metric, but it contained the kind of thing that was, I guess, not untouchable but very hard to make negative changes to, because it was tied to so many people's bonuses? Or is that, like --

A So I don't know this for sure, right, but I've had long conversations with people who have read these documents who are well-versed in Facebook where we are like, how the hell -- how the hell did Mark make this decision?

Like, the literature inside of the company has no ambiguity about what the impact of downstream MSI is. That's why they turned it off when the riots were happening in the United States, right, in 2020, for civic content, not broadly.

And the only thing I can imagine is, you know, like I said before, the management philosophy inside of Facebook is that you don't need to manage people, you just need to set, like, the right goalposts. Like, literally, my job for interviewing candidates, like, when we were interviewing them for jobs, was to assess, could they figure out what the -- could they figure out what -- like, let's say, I presented a product problem to you, like, some user problem. Could you articulate what the goal was to solve, and could you articulate what metrics would be appropriate to solve that problem? Right?

And the reason why I literally had a role in the interview process with just them is that a huge part of the management philosophy of Facebook is that you shouldn't have to tell people what to do, you should just pick the right goal metrics; and then, as long as they move those metrics, they're doing a good job.

Q And there was no consideration that, like, some other team -- there may be a very good reason why a particular metric was going down, so we're going to give you a handicap when calculating your bonus?

A I am not aware of that process. Right? Like, I'm not aware of a process where they would come in and say, hey, we know that we dropped MSI by half a percentage point, and, yes, the goal for the entire team across, like, the entire news feed team, across the quarter, was to move it by 0.8 percent or something, and, yes, we know we dropped it half a percentage point by doing this thing, but we're just going to give everyone a half-of-percentage-point addition, I am not aware of that process. It might exist, but I am not aware of it.

Q Okay.

And this is something you've mentioned a number of times, "civic content" you've called it. How does Facebook determine whether something is civic content, or how does Facebook assign content to various categories?

A Great question. This is one of my core complaints with Facebook from a democratic accountability perspective. Facebook does not disclose how it trains these classifiers. It doesn't say what examples count -- like, what's their policy? What counts as being in? What counts as being out?

My understanding of what counted as civic issues was -- I know that some of it was, like, directly political, like, a candidate or an election. I know that some of it was classified as social issues, and there was, like, a list of things that counted as social issues.

For context, I know that at some point, I think maybe in 2019, or maybe it was even 2018, if you had a civic page, or pages that was classified as civic, like, I think there was some kind of process where you had to get approved to run ads.

And to give you a sense of, like, how this concept of civicness, like, adds complexity, like, Ben & Jerry's often takes positions on, say, environmental issues, and that would be considered a civic -- like, I believe their account had to be, like, reauthorized or something to run ads. Like, they were now considered, like, a civic proponent.

I don't know what the classifier was like at the point that Jan 6th happened. I remember when I was working with this data -- so I worked with this data some in the summer of 2020. And I remember, at that time, they had four sub-buckets within their civic classifier. I don't remember what those four buckets were, though.

Q And --

A And it's important to remember that, if it's anything like the other classifiers inside of Facebook -- like, there are documents in the cache saying things like, Facebook only catches 3 to 5 percent of hate speech, it only catches 0.8 percent of violence-inciting content. Right? So, if it's anything like the other classifiers, yes, they said they turned it off for civic content, but they probably only turned it off for a tiny fraction of civic

content.

Q Can you just talk a little bit more about that? We were going to get to it later, but I think it makes sense to just --

A Sure.

Q -- focus on it now.

A So let's roll back our minds to when we were looking at the Jan 6 document.

Q We can go back to the January 6th document if you --

A No, no, we don't need to look at it. I just need you to remember back.

A number of the interventions were around -- instead of actioning in a step-wise way, acting in a proportional way, right? So saying things like, we'll look at how probable it is that something is violating, and we will -- we'll look at how probable something is as violating, and we'll, like, demote it proportionately to how likely it is that it's violating. Right?

For many things, like hate speech, there was not, like, a proportional intervention, or at least that's my understanding. There was -- either you were in or you were out. Right? And, in the case of hate speech, only 3 to 5 percent of hate speech was being taken down.

And a thing that's really important for people to understand is -- Facebook is going to come back and say, only -- I think the number they use is only 0.5 percent of content or 0.05 percent of content on Facebook is hate speech -- they're bad at detecting hate speech. And so, when they say that's how much hate speech there is, like, I don't think that's an accurate number.

But the other thing is, Facebook, whenever they launch one of these safety systems, they do an evaluation where -- you know, there's a concept in artificial intelligence called precision and recall. This might be true in more algorithms in general,

but you do it, at least, in any kind of retrieval system or detection system.

So, if you are trying to decide when to act, you can set your configurations to either be more precise so you'll have fewer false positives or you can set it to cover lots of things. So you have to come in and say, how much do I care about making sure bad things don't get through, and how much do I care about having it being absolutely right? And those things directly trade off. Like, if you want to catch more things, you have to accept more false positives.

Facebook's -- the sensitivity point that they would set things at was very, very high, which is, I think, okay, right? But it meant that they captured very, very few things.

And this is part of why having a strategy that's based on good and bad content I think will inevitably fail, is, like, you have to then accept that, like, you either are not going to get a very high fraction of the bad things or you're going to have a high false-positive rate and good things are going to get taken down sometimes.

And so that's what -- that's the context on that.

Q In this document, there's a table called "Summary of Proposals."

A Hmm. Okay. Let me scroll down to it.

Q The same colors up here -- red, green, and yellow. Keep going. There it is. This page and the next pages.

A Oh, interesting.

Q So can you talk -- I mean, what do you know about how this kind of information is filtered up and presented to Mark? Who makes decisions about which policies get flagged to him? Who makes decisions about how these colors work?

A Sure.

Q Et cetera.

A So I'm going to tell you my best understanding of this. I don't know if this is

actually how it was done, but if it was like other projects, this is how it would've been done.

So they did a broad call for suggestions of possible interventions. In this case, they have a bunch of these different ones here.

Oh, interesting. "Manufactured virality" is a really big problem. We can have a conversation on that later. It's one of the columns.

They did a broad call for suggestions, and the first winnowing stage was, you know, Anna Stepanov and probably her team looked through all those things, and, you know, it's possible that they came up with some different criteria and then scored all those items based on those criteria and then ranked them. That seems very, very likely.

So, often, what they do for those is they score them for, like, what the possible impact would be and, like, the difficulty to enact it, and then maybe, like, collateral damage. Right? And you can imagine using a system like that and then prioritizing what to focus on.

The next level of winnowing was they probably did a review with whoever was the team that she was under. So that might've been with Guy. It might've been with, like, maybe a set of leaders from Community Integrity. I don't know. At that stage, probably some things were -- I bet some colors in this chart --

[Audio interruption.]

A Hmm? Yes?

Q Sorry. Nothing.

A I'm guessing, at that point -- so it's probable that the team suggested colors initially, and then I'm guessing probably some of those colors were changed in that review. Like, they probably got feedback, and some things were changed.

And then that final version was probably what went to Mark.

Q And do you know what the M team is?

A The M team?

Q Yeah.

A I think that's the people who directly report to Mark.

Q Okay.

Is there anything else about this document that you think we should talk about before we move on?

A Sure. Give me 1 second.

God, some of these are so interesting.

So, if you look at number 6, "Manufactured virality," so manufactured virality is where people go and harvest content from other places. Like, they might look in a big group and see, like, what's the content that goes most viral, and then make a page where you just post that content because you know that you can growth-hack your page.

So that intervention -- so it says "Manufactured virality," it's number 6 intervention, it says -- so there's a row that says, "Tradeoff on value to people," and it says, "No expected decrease in sessions. Likes may decrease MSI somewhat." And I think that's one of the ones where it wasn't done in the end because it hit MSI. Because I think there's another document that talks about that.

Interesting.

Almost done.

So this is a great document in terms of, when you read across that row, what I find so interesting about this is, the data on MSI itself -- like, there's a number of documents in the cache that are reviews. Like, they went and analyzed MSI after it launched. And, allegedly, the reason Facebook did MSI is because they thought that it was important for people to interact with each other. Like, they shouldn't passively scroll; they should

interact with each other.

And yet, because the content is amplified and the system is biased towards is extreme content, when they asked people after MSI rolled out, is your news feed more meaningful, those studies had people saying, my news feed is less meaningful. Right?

And so they didn't do it to increase meaningfulness. They did it because content production was declining on the platform, and the only thing they found that could consistently increase content production was for people to get more likes, comments, reshares. Right? Those little hits of dopamine made people produce more content.

And what I find amazing about this chart is that they have a column that says, "Tradeoff value to people." And you can see that they rationalize why MSI is important in terms of, quote, it must be valuable to people, because every single one of these mentions, does it hit MSI? But if MSI has been shown -- it wasn't originally improving quality for people; it was just improving content production. It's funny how what is good for the company and what is good for people get intertwined.

I'm almost done.

Okay. Is there anything else I want to say on this doc? I'm almost done.

I think -- I think I'm done.

Q Okay.

As you've seen -- I just wanted to make sure everyone can hear me, because I've switched to a different microphone. As you've seen, we have a lot of documents that we'd like to ask you questions about, but we're kind of getting close to the end, and I want to make sure we have some time to step back, take some big-picture questions before we do that, and if we have time, then we'll jump back into some documents.

A Uh-huh.

Q So these questions are going to be just about Facebook, January 6th, the end

of 2020.

What was it like to work at Facebook in December and early January, before January 6th?

A So I was, thankfully, a little insulated, because I was working on counterespionage by that point, and the Threat Investigator org was more insulated than other parts of the company or, like, other parts of Civic Integrity. Like, my life didn't change that much when the reorg happened. Like, I had a new boss, and I had -- I had a new boss, and I had -- but, beyond that, my life didn't change that much.

I think there was a lot of skepticism about Facebook's seriousness, because, like, when I left, just so you guys are aware, when I left in May of 2021, the pod that I was in -- so, like, there were maybe seven product managers and program managers in that pod -- some of it was related to civics ban, some of it was information operations, some of it was coordinated inauthentic behavior -- every signal product manager or program manager quit within a 6-week period. So I think the end of 2020 was kind of like the moment where we were all still holding our breath, like, we were seeing, like, does Facebook still care?

And the fact that -- you know, my pod was maybe 75 percent of people who had come from Civic Integrity, right? Like, we had gotten reorganized into this pod. And, in December 2020, I think people were like, well, let's see if this actually -- like, are we still making a difference?

And so, yeah, I think it's always really disruptive when there's a reorg. I think there's even some discussion in various documents in the cache on this. Like, when there's a reorg, you lose a quarter, you lose half a quarter, just because it causes lots of chaos, right? Like, every sub-team inside of Facebook operates slightly differently, and so you have to be able to go in there and, like, figure out how to succeed in that new

system.

And that definitely happened in the wake of the reorg. Right? Like, you know, you have to go and reestablish your bearings and reestablish relationships every time there's a reorg.

Q Do you think the staff was exhausted?

A Yes, unquestionably. The staff was exhausted before the 2020 election.

Like, I don't know if I have -- I think I have at least one document on this, which is, like, what was -- they used to do these internal sentiment surveys. And this might be an interesting thing for you guys to subpoena from Facebook. They used to do these things either once a quarter or once a half, I'm not sure, which were basically, like, employee happiness surveys. And, you know, they would ask questions sometimes like, do you think Facebook is doing work that's good for the world, that kind of thing.

And, consistently, the Civic Integrity org had substantially worse scores than all the other orgs. And part of that was that we were ridiculously understaffed. Part of this was that we didn't have institutional support. Part of it was it's really, really damaging to a person to know secrets that potentially are going to kill people and then not to be able to do anything on it.

And so people were definitely exhausted in December 2020, and they were exhausted, like, even before that.

Q Do you know if anyone at Facebook was paying attention to the "Stop the Steal" movement?

A I believe we have a document that outlines what was done on "Stop the Steal." That is the only information I know about what was the state of "Stop the Steal."

Q Did Facebook do a good job of foiling foreign interference in the election?

A You know, I know that they did stuff. Given what I saw in terms of how

understaffed things were, I can't imagine they did an adequate job of deterring foreign interference. I can't imagine that it was an adequate job.

And Facebook claims that they had 200 people working on information operations. I have no idea where those 200 people would be. Because, at least in the Threat Intelligence org, I think when I left, there was 17 people on information operations. Right?

And -- oh. This is another, like, bad one. So, in 2016, one of the major drivers of misinformation was the Macedonians. Like, the Macedonians came in and had this whole cottage industry around driving traffic to these fake news sites based on, like, recommender systems at Facebook. And so there is a model inside the company for, like, thinking about where do these coordinated systems come from, like, coordinated efforts to spread disinformation, misinformation. Some of them are political. Some of them are commercial, right? People are trying to make money.

The eCrime team was shifted onto looking at those kinds of operations, kind of like what the Macedonians did, for, I think, a couple of quarters before 2020, because it was discovered to be such a mess. I can't imagine that team was able to do as good a job as, like, the public deserves, because there were only, like, six people on that team. Like, it was very small.

So, yeah, like, I can't imagine that there was an adequate job done on -- oh, actually, important thing for you guys to know: So, like I mentioned earlier, I did across -- across I-3 projects related to proactive detection. To give you guys context on why that matters, my team, which was counterespionage, did not work on proactive detection systems, because at any given time we only had enough staffing to work on a third of the cases that we knew about. So it was seen as there was no point in making proactive detection systems because we already couldn't handle the cases that we had.

So the metaphor I usually use for this is, we knew we were interacting with an onion, we knew we were only interacting with the outer layer of the onion, but we had no idea if the onion was the size of a baseball or a beach ball. And the same was true for information operations, right? And the same was true for these commercial operations. There was very, very minimal proactive detection systems because there wasn't enough staffing to handle what already existed.

And Facebook had made a public statement, Mark made a public statement to Congress that, when they found information operations, they would take them down and inform the public. And so there was a hesitancy to detect too much, because there wasn't staffing to handle what was found.

Q So it was like an ostrich.

A Yeah.

Q If you could say that you weren't convinced that it existed --

A Exactly.

Q -- then you wouldn't make Mark Zuckerberg a liar?

A Exactly.

Q Okay.

So I have a question about resourcing. So there's resourcing on your team, on the team you were working on, where you were doing all of your work. But weren't you dependent on other teams to do some work that you recommended as well?

A So, when I worked on -- when I worked on counterespionage, our -- yeah, actually, that's true. Of course we were.

So I'll give you an example. I worked with the malware team, right? So there was a team that had a system for detecting malware. And in the run-up to end-to-end encryption, my team was extremely concerned about the fact that we were going

to -- like, at the time, a thing that our team did was that, if we detected you were being sent malware by, like, a foreign agent, we would break that malware, like, we would put a little change into the file if it was, like, an Android APK such that, when you tried to install it, it wouldn't successfully install. But we would warn people, "This is malware," because we could scan the attachments.

And one of the things that we wanted this other team to do was to give people at least the choice. Like, once we moved to this end-to-end encryption world, people should be allowed to choose, do they want their attachments scanned for malware? Right?

And, like, that was -- because everyone is stretched, like, you often can't get things from other teams that you need done. And so, like, we couldn't get a commitment on when that would be done. And I don't think it's become a problem, because I don't think they've -- they haven't launched default end-to-end encryption yet, right?

But that's the kind of thing where, like, yeah, we definitely had things where we needed other teams to do things.

Q So, on the day of January 6th, what stands out most in your memory?

A In my memory? So I was in -- I was visiting a friend in Nevada City, and I was not watching Jan 6th very closely, because, like, my team mostly dealt with foreign espionage, right? And I remember checking to see, was anything being done? Because I was watching the news on it, at least. And that's how I got these photos, right, is I took pictures of them on that day.

I was not directly involved in those actions, though, and so it was not a particularly different day for me.

Q Did you see anyone who you thought behaved admirably in the days after January 6th?

A Because I was working remotely and the only thing I saw was what happened in these docs, I don't have any -- I don't have any first-hand experience of that.

Q Okay. And so I guess the counter-question is, did you see anyone who you thought did a particularly bad job on the day of January 6th or immediately afterwards?

A No, because I wasn't working directly on it.

Q So what would've been different on January 6th, in your opinion, if Facebook had not dissolved Civic Integrity?

A I think -- so you know where we saw the difference between how much progress was manned, that doc between Jan 6th and Jan 7th? I feel confident that, had Civic Integrity not being dissolved, there would've been a war room already stood up on the 1st of January, on the 25th of December. Right?

Actually, another thing that I think is important to note is, in Silicon Valley, people often take off the week after New Year's. And it's very possible that a thing that made us even worse was that people left on vacation on, say, the 19th, and people weren't back from vacation yet, or not enough people were back from vacation.

But in a world where Civic Integrity had not been dissolved, I think this would've been a tracked issue on the 15th of December. Like, someone would've been watching, had they not dissolved Civic Integrity.

Q Okay. So that seems like stuff that happened on or around the actual attack. In terms of the planning and the organizing and the messaging and the misinformation that appeared on Facebook, do you think any of that would've been different if Civic Integrity hadn't been dissolved?

A Yeah. Because, like, you know, what we were talking about before about this idea of, like, the war room would've been in place, the war room would've already started taking actions before Jan 6th. They probably would've started taking actions,

you know, 2 weeks or 3 weeks before Jan 6th.

And, you know, when they did the reorg, you know, it makes people not productive for a period of time, because, like, everyone is trying to find their bearings. We would've gone through -- started going through the planning process for the new half of the new quarter. And I remember even that got delayed because, like, people were so disorganized in the advent of the reorg.

And so, yeah, like, there would've been less misinfo, there would've been less violence-inciting content, all these things, because a lot of the interventions that got turned on on Jan 7th would've gotten turned on 2 weeks earlier.

Q Okay.

Now, putting aside the reorg, think broad, think narrow, think -- take this question however you want. But what was the biggest mistake Facebook made regarding January 6th?

A Great question.

So Facebook knew there were problems with groups. And I think groups were a major, major driver in all the problems that happened on January 6th. Facebook knew there were major problems with groups, in terms of group recommendations, in terms of this idea that you can invite someone to a group, you invite thousands of people to a group. Like, I think the top "Stop the Steal" inviter -- or maybe it was QAnon. It was either QAnon or "Stop the Steal," the top inviter invited 300,000 people. Like, one account invited 300,000 people to these kinds of groups.

And, like, Facebook knew that there were these problems around, like, rate limits, around -- like, there's documents in there, and this has been written up before in the public sphere, the idea that Facebook's detection systems for good and bad groups take a long time to evaluate groups. It's like those systems run on, like, a 2-week delay. And

people have raised questions before on, should we recommend a group when it's less than, say, 2 weeks old? Because if we can't tell if it's a good or bad group yet, like, we're basically flying blind.

And, in a world where, you know, you could have -- let's imagine you had a group where there were 20 organizers on that group, and those 20 organizers might actually be one person with 20 accounts, right? If each one could invite 2,000 people to the group, you know, that means you can invite 40,000 people every day across those 20 accounts, right? Very rapidly you can have a huge distribution network. And if they don't have to actually accept that being a member in order to get their content in the feed, you can see why this becomes a huge, huge, huge vulnerability.

Facebook knew all these things months and months and months or years before Jan 6th happened, and yet they were fine with it, because, when you made more assertive safety settings, you know, you made groups' growth lower, right?

And, in a world where Facebook doesn't, like -- I think an important thing that is not obvious to people is, let's imagine a world where Facebook was just about your family and friends. The amount of time you would spend on Facebook is substantially less than you spend today, right, or the average person spends today. Because Facebook has figured out that they can put way, way more content in your feed if they push you to join groups, especially large groups.

And, at the same time, Facebook has a financial incentive to try to get you to stay on the site as long as possible, because the longer you're on there, the more ads you consume, right?

And Facebook has made these tradeoffs consistently such that they protect the growth of groups, they protect distribution of content from large groups, because they have an outsized impact on the total amount of content consumed on the site. And

they have prioritized that configuration over known safety problems for a very long period of time.

So I'd say that's the biggest, like, set of mistakes that Facebook made.

Any other specific things? I think the fact that there's not enough, like, ongoing monitoring. Right? Like, there should be -- the fact that there needed to be a war room set up in order to have these kinds of functions get flagged. Like, why was it dependent on having Civic Integrity watching? Like, why weren't there other safety systems? That seems like a big negligent thing.

I think that's all that's top of mind for me.

Q Okay.

Group integrity, where did that sit?

A So group integrity -- so I think it's always important to think about, when you're doing structures of things, is each major product family also likely had a sub -- like, a specific surface area integrity team.

So there's a guy named Nir, N-i-r, and he was on some of those Jan 6th break-the-glass measures. Like, he was listed as the responsible person. So I believe Nir was the head of the groups-specific integrity team.

And then you would also then have, like, a team that worked on, like, a problem space, so maybe it's, like, violence incitement. And they might do group-specific interventions, right?

So the responsibility for groups integrity as a whole was spread between specific problem teams and specific surface teams.

Q So they worked together.

A They worked together, yeah.

Q Okay.

A It's all a question of, do you follow a single problem across multiple surfaces, or do you only deal with a variety of problems that occur on a specific surface? And so Nir was responsible for integrity problems that appeared on groups.

Q And so my understanding is that the focus on groups, a lot of it traces back to a Mark Zuckerberg decision to focus on groups. Is that your understanding? Or is that --

A I don't know if it's a Mark-specific thing. It might be a thing where, you know, someone pitched to Mark and said, hey, groups -- I know a thing that I was definitely told was that, like, your participation in groups was directly correlative with the likelihood you were going to still be on the site 30 days later.

And so I think -- I don't know if it was a strategy that came from Mark or if it was a strategy that came from some growth-hacking person. But it was definitely a philosophy within the company.

Q Okay.

[redacted] I've taken the microphone for a long time. [redacted] I'm happy to keep going. If you have some specific questions that you want to follow up on from earlier, that's okay too.

[redacted] I do. And I'm happy to jump in here, unless there's something else you definitely want to get on the record before we wrap.

[redacted] Oh, I have some other things to ask about, but I wanted to make sure you had some time.

[redacted] Oh, great. Well, thanks.

I will try to do some rapid-fire through this, Frances, going as quickly as possible while still giving you time to answer. And then really just a few followup questions that I've written down from your conversation with [redacted] over the past few hours.

And forgive me if you've covered any of this. For about the past 20 minutes or so, I was dealing with an emergency that came up with the team, so I've missed the last piece of what you've said. So, if you've already been asked this question in the last 20 minutes, just feel free to say that.

Ms. Haugen. Okay.

EXAMINATION

BY [redacted]

Q The first: You said early on in the interview that you joined Facebook because you lost a friend to radicalization online. And then you said that, when you joined, I think you said something along the lines of it became immediately apparent to you that things were far worse than you realized and maybe couldn't be fixed.

And I just wanted to unpack that a little bit and try to understand what your thoughts were before you came to Facebook and what happened when you first got there that made you realize that it was worse than you thought.

A So a thing that Americans don't understand, across the board, is, they are not aware that the version of Facebook they see every day is the best version of Facebook in the world. Right?

Like, I joined thinking I was going to work on the U.S. 2020 election, and I got put on international misinfo. And, internationally, misinfo is way, way worse than in the United States, because people are dying. Like, the stuff with Myanmar had not started yet, but it was, like, a known -- it was known that misinformation was actively fanning ethnic conflict in Africa, right, or in Southeast Asia, right?

And I had never thought about that before, right? Like, I am very fortunate to live in a beautiful, relatively stable place in the world, and, you know, most of my friends are privileged and, you know, use Facebook in a very pleasant way. And I had never

thought about the idea before that, like -- I think this is actually important for folks to understand in general.

Like, the version of Facebook that probably everyone on this call interacts with is a substantially nicer version of Facebook than what the average person in the United States interacts with. That's because the average person on this call has more privileged friends than the average person in the United States. And that translates into people that produce more original content. It translates into people who know how to have arguments in respectful ways, right? Probably less conspiracy theories, on average, or less, like, flagrant misinformation.

And, you know, we just don't think about it. We don't think about this idea that there's not one Facebook; there's many, many Facebooks.

And Facebook goes to great lengths to make sure that people who have influence see a version of Facebook that doesn't inspire them to act. And that means things like the whole cross-check program -- the whole reason why cross-check exists is they don't want people who have the influence to put pressure on Facebook to be annoyed at Facebook. Right?

And I had no clue that any of this stuff was happening until I joined.

Q So is there any document that would show what you just mentioned about the people on this call having a nicer Facebook than most people in America? Is there anything that would demonstrate what most people in America are getting?

A I think probably the closest thing to that would be, some of the COVID documents talk about -- so a thing that I worked on as part of the narrow cached misinformation under civic misinformation was -- and this is a thing you guys could subpoena from them, right -- is they segmented the United States into around 600 sub-communities. And this is based on people's pages, groups, interactions, all those

things.

And the reality is that all harms on Facebook are not evenly distributed, right? So, in the case of the misinformation, a very small fraction of people get overexposed to misinformation. In the case of COVID, 4 percent of the population gets 80 percent of all the COVID misinformation.

Q Hmm.

A And I'm sure the same is true for violence-inciting content. I'm sure the same is true for hate speech. There's a subpopulation that creates it and a subpopulation that has to consume it. Right?

So that data exists in the world. I don't think I have a document that says that.

Q Okay. They have something -- Facebook should have something that speaks to this.

A I could tell you how to do the data request for it.

Q That would be fantastic. We can follow up on that with Libby.

A Yeah. And there's documentation in the cache -- I don't know if you've received it or not -- that explains how segmentation works. And you guys can read through that, and we can talk about, like, how would you do that data request.

Q Got it. I think we do have that. I think [redacted] read that. I don't think I have yet. But that's really helpful.

Next question -- and, again, you might have covered this. But I know we talked about Civic Integrity being dissolved in December, but do you recall when in December it was dissolved?

A I'm pretty sure it was the first week in December.

Q Oh, so very early. Okay. And you may be aware that Facebook is claiming that that's not true -- not just that, you know, Civic Integrity was dissolved and, like,

reprioritized in other sections, but that it didn't actually get dissolved until sometime in 2021. Have you heard that argument, and what's your response on that?

A I can't imagine them.

Q Yeah.

A The -- can I have a sidebar with my attorneys? Is there a way for us to -- can I sign off?

Q I think you should put yourself on mute and turn off your video and probably give them a phone call.

A Okay. Give me 1 second.

Ms. Liu. I'll start a call.

[Discussion off the record.]

Ms. Haugen. Okay. So, with regard to proof of when they dissolved Civic Integrity and how they described what they did to the population, it is likely that even next week I can be helpful with that, but unfortunately I can't give you more details at this time.

BY [redacted]

Q Okay. Great. We'll put a pin in it then. That's good to know.

A Yeah.

Q You also mentioned early on that Civic Integrity was dissolved because -- or you think it was dissolved because it, I think the way you put it was, found problems and opened cans of worms that were not easy to close.

A Uh-huh.

Q And I think you've talked about what some of those problems are in a way that's really helpful. But I'm wondering if there are other teams outside of Civic Integrity that were also finding problems and opening cans of worms that you think we

should be looking at.

A Interesting.

I don't think anyone was as disruptive as Civic Integrity. Because, if you think about it, like, the scope of problems that Civic Integrity was responsible for was things like genocide. Right? Like, in terms of things that Facebook considered, like, real existential issues in terms of intervention -- actually, a thing that I don't have documentation on but I was told at an all-hands was that Civic Integrity's budget came out of the antitrust budget. Right? Like, it was not in place for altruistic reasons. It was in place because Facebook understood that, if they really messed up the 2020 election, if they caused another genocide, it was likely that something would happen, right?

And I don't think anything was on -- any other team at Facebook had -- like, because I think it's good for us to touch on really briefly, what is the lifecycle of solving a problem? So, like, the lifecycle of solving a problem begins with things like researchers going and talking to people in the field, right, or involves data scientists identifying new ways of detecting something.

And, often, kind of, like, the chronology would be: A researcher discovers a problem by talking to people, or it's in the literature or something. Then maybe a data scientist figures out a way to detect it. Then maybe a data scientist or an engineer works on -- and maybe works with a product manager to figure out, like, how would you describe solving that problem?

Then a product manager would work on articulating exactly what would happen and on getting all the approvals and -- like, other teams you would have to cooperate with in order to get that passed and approved, like, Policy, for example, External Relations, PR, Communications.

Then engineers would build it, and then you'd launch it. And then people, like, in operations would monitor for what the effect of that is. Like, did people start complaining about something?

And the only place in the company that was doing real initial research on things like ethnic violence was Civic Integrity.

Q Got it. Okay. That's really helpful.

Also early in the conversation, [redacted] asked if you thought that Civic Integrity should've been dissolved and you said no, and you said that was because there were other things that could've been done to make the team happier and more functional other than splitting them up.

A Yes.

Q So I was hoping to unpack that a little bit. I mean, what do you mean when you say that the team wasn't happy or wasn't functional? What was going wrong with the team that you thought needed to be fixed? And what did you think they should've done instead of splitting it up?

A So I fully acknowledge the team had worse, like, happiness scores, right, than other teams, or, like, you know, it was viewed as being less impactful. But, like, part of why it was viewed as being less impactful is, like, we have a way larger remit of problems that we were solving. So, like, they would have a big problem, like voter disenfranchisement, and they'd put, like, one engineer and one data scientist on it. And you ended up in these situations where, like, it was like they were trying to set up an organization as large as the entire rest of the integrity org but they were trying to do it in, like, 2 years with a tenth as many people, right, or a fifth as many people.

And if they had staffed more data scientists on Civic Integrity, they would've come up with more metrics to show our progress, and we would've operated in a way that

looked like success in the language of Facebook. But because we didn't have enough data scientists, you had a bunch of these problem areas that didn't have, like, defined metrics. And so the work that we did was not viewed as existing, because, like, at Facebook, if it's not measured, it's not real, right?

Or, for example, a lot of people were frustrated because they would come up with solutions and, because those solutions would hit MSI at all or would hit growth at all, they wouldn't get passed.

So, like, is the solution to get rid of the team? Or is it to have someone step in and say, "Guess what, the system as a whole has problems, and we need to do something"?

And so I think they looked at it and said, this team has some of the lowest happiness scores at Facebook, we don't view them as successful, and they create liabilities for the company, we're going to dissolve them.

Q Got it. That makes a lot of sense. Thanks.

You started to unpack this a little bit with [redacted] but you mentioned a few times that, if someone like Samidh or Civic Integrity had existed in the run-up to January 6th, you think that Facebook would've taken different actions, and part of that is that you think they would've turned on some of these break-the-glass measures or identified some of these 7,000 people well before January 6th.

And I'm just trying to imagine, like, in my head what Facebook is likely to say in response to that. And I could see them saying, you know, this stuff was happening on our platform but, like, we had no indication that it was going to move from general misinformation or general talk to something in the real world.

And so I'm wondering if you could describe a little bit more why, not just in the instance of hindsight, but what do you think would've triggered Civic Integrity or Samidh

to do more beforehand that would've drawn the line between just some stuff happening on Facebook to, oh, we need to take some action to prevent something happening on January 6th?

A I think it's good for us to look at Myanmar as, like, a great analogy. Like, how did Myanmar happen, and what did they do differently after Myanmar? Or what did the United Nations tell Facebook it was supposed to do differently after Myanmar?

So Myanmar got as bad as it did because there was only one person in the company that spoke Burmese, and when that person tried to escalate that there were problems, their feedback never percolated upwards. Right? They were, like, kept -- that feedback was kept down, partially because there's a culture of not stressing negative things, right?

Stressing growth and optimism and focusing on -- like, Bosworth has said before -- he's, like, the CTO now -- that, like -- and this has been publicly reported. I don't have this document, but other people have it -- saying, like, it doesn't matter if we cause harm, because, de facto, the value of connecting people is more valuable than any -- like, even if there's suicide on the platform, whatever, it doesn't matter, because connecting people is more important. Right?

So people who toe that company line get rewarded. People who don't don't.

One of the things that you unflagged was that not only were there not people inside the company that spoke these languages; it was also that when people on the outside who saw what was happening tried to contact Facebook, they had no way of getting in touch with anyone significant at Facebook.

And they started a program called -- I don't know what the inbox is called. There's a special inbox. There's a dual messaging system at Facebook. So there's the messaging system that most people have for interacting with, like, comms and support,

and there's a special one that you have to be whitelisted into. And, initially, it was only in Myanmar so that NGOs could report to Facebook, like, when they saw stuff that was likely to cause ethnic violence.

And I think what happened with Jan 6th -- with Ethiopia -- so let's contrast what happened with Ethiopia versus Jan 6th.

With Ethiopia, there was a group inside of Strategic Response from January 2020 that was actively watching Ethiopia, right? In other situations, like, various national elections, Civic Integrity would actively watch how a situation evolved in that country. This happened over and over again. It happened with the EU parliamentary elections, the Indian elections, U.S. elections. They had a playbook for assessing risk in advance, having a war room stood up that would monitor things, or, at a minimum, having, like, priority groups that were watching.

And I cannot imagine that the hand-off -- so one of the things to remember, this is, like, an organizational structure inside of Facebook thing. You know we talked before about there's difference groups. So country operations has a budgeting cycle that is quarterly. So whatever those humans are focusing on, it is budgeted out in hours, and there's a quarterly process for allocating where their time goes. So, if you were going to have a group of people monitoring a situation that were not engineers in a war room, you would need to have them likely allocated through country operations.

Remember how I talked about before that, when the reorg happened, it threw off the planning cycle sufficiently that they delayed the entire planning cycle for Community Integrity by, like, 6 weeks. Like, I don't think we had to have our H1 plans set until the middle of February, because things were so disorganized after they dissolved Civic Integrity.

I can totally imagine that, coming into December, when country operations

would've had to allocate people -- I can imagine a couple different things. One is, it is likely that, going into H2 2020, so that's July through December, that when the staffing allocations were made for those humans watching all these alert cues, I'm guessing they probably didn't allocate as many hours after the election to before the election. And so, if you didn't have people in place who were watching the situation and able to focus on it and say, like, "Hey, actually, this is escalating more than we expected, we need more people involved," I can totally imagine that the people who would've made that call, had Civic Integrity been left alone, were distracted in December 2020, because their team had just been dissolved and they were, like, having to make new relationships with new managers and figure out priorities and everything.

And so I think there was almost certainly not enough people paying attention and watching. And that's why there was such a giant difference in interventions between Jan 6th and Jan 7th.

Q I know Civic Integrity was dissolved in December, but we started to see some of this ramp up between November 3rd and early December. Do you know if there was any attention being paid or any concern voiced about that ramp-up during the time period where Civic Integrity was still in place?

A I was working on -- you know, it'd be interesting. I have -- can you give me 1 second? I want to -- eh, I don't think I have a fast way to check for that. I might have -- I might have -- yeah, I'm not aware of anything off the top of my head. If Libby can take a note, I can go look into something.

I'm not aware of discussions around that. I know that people were really exhausted by the time the 2020 election had passed. And so it was another one of those things where, because things were so understaffed and people were, like, just running on fumes, I think that's another thing where probably not enough attention was

happening. I just don't know.

[3:31 p.m.]

BY [redacted]

Q Got it. And if they had been monitoring that stuff in November or December, how would they have been measuring that to know, like, oh, this is a problem that's escalating and we need to do something about it?

A So, for example, you know how I talked before about the idea of segmentation?

Q Yes.

A Like, you could imagine them having a system -- and I think they should have to have this system. Like, they should be tracking integrity problems at the level of segments. And it would be easy to see: Oh, interesting, there is an escalating amount of hate speech, there's an escalating amount of violence incitement amongst 1 percent of the population.

That is a doable question, and that's a doable thing to track. And I think that was just never a thing that was considered in the, like, repertoire of things to do.

Q Got it. So it sounds like, if they had been doing all of the ideal things, they would have seen this coming. But there's nothing that they routinely did --

A Because --

Q -- that they had -- sorry. Go ahead.

A Because things are so understaffed --

Q Uh-huh.

A So, as of December 2019, I was telling Samidh that you could be doing these things --

Q Got it.

A -- a full year in advance.

Q Okay.

A But, at the same time, the team that built that -- like, ended up finishing that system -- because I got taken off that team -- they were really overstretched, right? And we can go look at what they were working on in 2020, because I have a couple -- I know I have a couple documents from them. They may have been working on COVID, for example.

But it's one of these things where, if we had been at Google, there would've been literally 10 times as many people working on these problems.

Q Uh-huh.

A And so it's this thing of, like, yes, they didn't routinely do these things, but they were totally doable, and if they'd had more people, they would've been done.

Q Got it.

You also mentioned a whitelisting program in Myanmar where NGOs could send up a flag if they were concerned. Was there anything like that for the United States? And if people had raised flags about January 6th, where would we look to find documentation of that?

A So there was that program for the United States but, to the best of my knowledge, only for voter disenfranchisement.

Q Okay.

A And so this is another great example of, you know, if Facebook was doing a rigorous job -- like, think about that for just a second. So, like, the fact that the United Nations could come in and say, hey, you have a problem that, when the system goes out of bounds, you don't have a good way for NGOs to tell you -- like, you would think that, once that had happened once and you one genocide, you would think about that more

systematically for how could it apply elsewhere in the world. Right? I mean, the U.N. doesn't write these reports for their own good, right?

Q Right.

A But because things are so understaffed at Facebook, like, just that project alone, like, the thing that allowed you to, like, be an NGO and tell, you know, Facebook a genocide is happening -- when I'd -- I had that project for a little bit in 2019, right? Like, it was my project for months before it was decided that I had too many things on my plate and we needed to, like, have fewer -- we needed to divide it up.

That system was so undersupported, Facebook -- so, when you write software, you have frameworks that allow you to write software faster. And Facebook had gone through and written a new framework for writing their front ends, and you were no longer allowed to ship changes on software that didn't use the new framework.

And the system for NGOs to tell Facebook that genocides were happening was so underinvested in that we couldn't do even minor changes to it, because it was written in the old system and we didn't have enough resources to put it in the new system. Right?

And that, I think, got remedied after it wasn't my project anymore, but it shows you how little systematic thinking or, like -- like, Civic Integrity tried to be systematic, but they were so understaffed all the time that they never got to do what I think would've been an appropriate level of intervention.

Q Got it.

You know, I want to shift to after January 6th. Do you know if there was any postmortem or after-action report done at Facebook about the company's role in the lead-up to and what happened on January 6th?

A I would be entirely unsurprised if that happened. Like, I would expect that document to exist.

Q Okay.

A And if you can't find it, I would ask aggressively about whether or not it'd been deleted, right? Because Facebook --

Q So the indications that we've gotten is that there had been, sort of, team-by-team looks but no comprehensive, company-wide look. Does that sound right to you, or do you think there probably was a company-wide look?

A I can't imagine there wasn't a company-wide look, right? Like -- and if there wasn't a company-wide look, that alone is negligent.

Q Got it.

And, since you mentioned this, actually, the thing I was going to ask about next was, you mentioned that there was a pattern of deleting things that makes the company look bad.

A Yes.

Q So I was wondering if you could say a little bit more about what you mean about that. Like, who was deleting things? Who was making the decision to delete things? Was that centralized or dispersed among the team? Just elaborate a little bit more on the pattern that you talked about.

A I don't know who made the call. So the first person I noticed this for was the woman who all of her notebooks had been deleted.

Q Right.

A And that wasn't a generalized pattern. Like, I had never seen before, like, oh, all of the content from someone is gone from, like, over -- like, after they left.

The second person I encountered it for was, I had a colleague that I had worked on the civic influence modeling stuff, and he had uncovered a bunch of stuff around, they're called human-driven fake accounts. They're not algorithmically driven. They're

not truly a robot. They're, like, a kid in Pakistan, because there's a whole cottage industry there, or in Vietnam, certain parts of Africa.

It's a very geographically isolated cottage industry, where people go and play with an account for a month, and this kid makes, like, a dollar for having, like, made a profile and interacted with stuff and pretended to be a person for a month. Because most of the detection systems for finding fake accounts only watch for a certain period of time, and if you can stay real-enough-looking for long enough, you get basically blessed as, like, you are a real person.

And, like, he had raised to his vice president, like, to -- the way he described it to me was, he talked to the vice president for integrity -- that might've been Guy Rosen, it might've been someone else -- and said, like, hey, we need a red team.

So, in security, there's the concept of the blue team and the red team, like, when you do, like, a war game, where the blue team is the good guys and the red team is the bad guys. And you basically, like, think of, like, how would you break Facebook? And he was like, we need an integrity red team, because, like, I know how you would take down Facebook if you wanted to take down Facebook. Like, this is a huge problem.

And, literally, the next day, he got locked out of all of the Facebook services. His manager didn't even know what had happened. His manager got access reinstated the next day, and, when he logged in, all of his documents were deleted. All the jobs he had running, like, all the computer programs he had running -- because sometimes you run jobs that take days to run, right, if you're doing AI or things like that. All the jobs he was running were gone. All of his data was deleted.

And, like, I don't know how widespread that is, but, given that even I saw it twice, I assume it happened more than twice.

Q Got it. Okay.

But it doesn't sound like there is any other particular documents that you think might have been deleted that we should be looking for. It sounds like this was just something that you noticed as an anecdotal thing that you think is broader but there's not anything in particular we should be looking for?

A I think it'd be interesting to just look through the documents in my disclosure and see --

Q Yeah.

A -- which have been, like -- like, re-request them, right? Just go through and re-request everything that's in the disclosure, and if they can't provide something, then that's a red flag.

Q So, still on the issue of deletion, but slightly different, and I'm not sure if this is outside of your wheelhouse, so feel free to say if it is. But you may have seen, there was reports of a lawsuit in New Mexico about a militia out there. And, as part of that lawsuit, the prosecutor has been looking to get documents from Facebook about the militia group that -- I think it said New Mexico -- I can't remember exactly what it's called -- Civil Guard or something along those lines.

But the issue is that they're trying to get documents about who were members of that group and metadata about that and content that was deleted, but Facebook took that group down as part of a content moderation procedure. And they have apparently represented to the prosecutor that they don't have any metadata or content from any of those users because they routinely delete it as part of content moderation.

Does that sound right to you? And, if so, how do they address or prevent recidivism if they're constantly mass-deleting information from the groups that they're deleting?

A Great question. Good job. That is a bad one. That is a huge problem,

right? It's a huge, huge problem. Because when they delete accounts, it ends up in this problem where it's hard to do recidivism protection. Like, this is a known thing across a bunch of harm types in threat integrity, like, the threat researchers, I-3.

I would be unsurprised if they had trouble doing that. This is the kind of thing where I strongly believe there needs to be more oversight of Facebook, because a lot of decisions that they make are very convenient for them, right? It makes it hard to actually hold them accountable. And the public doesn't get a say in how these systems work, and, as a result, like, they don't get justice.

So I can totally imagine that that is true. In general, they delete all the content -- so one of the consequences of that process is that it makes it very difficult to train the eyes to find these bad actors, because there is no training data, right? Like, Facebook deletes all the -- I think they delete all the logs they can after 90 days. And they claim they do this for GDPR, and I don't think that -- I think that might be technically true, but I think it is also extremely convenient.

Q Got it.

And I think you mentioned earlier that you did do some work on issues of recidivism. Did I hear that right?

A Uh-huh.

Q And did this issue come up at all, in terms of how do you train a computer or how do you train people to identify groups that are popping back up under different names if you've deleted all their stuff? Did that come up in your line of work at all?

A Absolutely. And this is particularly a problem for information operations, because, in the case of espionage, the most productive thing for us was to watch accounts over time and then act after a bit. Because it was better to help the people being targeted in those operations than to necessarily take the accounts down.

Because, once you had them, you could watch them, but if you lost them, it'd be harder to go and, like, re-find them, right?

In the case of information operations, because Mark promised Congress that when they found something they would take it down and inform the public, it meant that it was actually hard to accumulate training data to find other bad actors, because you couldn't watch them over time or you couldn't watch them for a bit and be like, let's learn their patterns or whatever and then take them down. Because, like, you did to take them down immediately upon finding them. And because data in the system would get deleted, like, then you wouldn't have stuff to act on.

Q Do you know what would've happened to content that was moderated on January 6th in real-time? Would that have gotten auto-deleted? Or do they save that in, like, purgatory before it gets deleted? What's the process there?

A I don't know. I would guess it would stay in the logs for 90 days and then get deleted.

Q And this 90-day deletion policy, what exactly does that apply to? Is that all content moderation, or is that some other subset of information?

A That's just the logs.

Q Okay.

A So, like, everything that happens on the site, you can query it, right? And the -- you can query it, but, at the same time, you know, if you get more than 90 days out, unless you save that data in some intentional way, all those logs automatically get deleted. So it means that, by the time you, say, file a request for something, it's too late. Because, usually, by the time we realize there's a problem and we go through, like, mechanisms of justice, more than 90 days have passed.

And so it's this really interesting question of, like, should that number be 6

months? Right? Because even 6 months would allow for substantially more safety.

Q Yeah.

I know we only have 15 minutes left, and I want to give [redacted] the last 10 or so, so I am going to try to ask a couple -- I could probably talk to you for a few more hours. I'm just going to ask you a couple of more quick questions so I can pass the mike back to [redacted]

But I did want to try to get a better sense of MSI.

A Sure.

Q I'm wondering, does that get measured by kinds of content? So, for example, could we get measurement on how much QAnon content was shared via MSI or how that scored or how "Stop the Steal" content scored on MSI? Is it broken down in that way at all, in terms of content? Or is it just more general than that?

A You could go -- so some things you could do are, you could go in and say, okay, show us content that is QAnon-related, and now add up all the MSI that that content generated. So you could go in and say, let's count all the likes, all the comments, all the reshares; let's use the weights that are in the standard measure for MSI -- right? Because, like, a reshare is worth different amounts than a like, that kind of thing. You could then compare that amount of MSI on that content on, like, a per-item basis to the average amount of MSI on content in general on the site.

That's an analysis you could do.

Q Okay.

A You could go and similarly -- so, like, a different question is around what parts of the scoring system led to that content being distributed, right? So you could go on there and say, show me the scores on the QAnon content, when they were shown in people's news feeds -- and that's another thing they can get for you, right? There are logs that show the exact scoring parameters that were used on each piece of content.

And you could say, which factors in the scoring were higher, like, disproportionately higher, on QAnon content than other kinds of content? And that would let you know what, like, downstream MSI unquestionably plays a larger role for QAnon content than other kinds of content.

Q Got it.

And with my last 2 minutes -- I do still have, like, two more pages of questions, but I'm going to ask this one and then pass it to [redacted]

You mentioned a study in Germany that said that 65 percent of people who joined neo-Nazi groups joined because Facebook recommended them. Are you aware of any similar studies about the United States, one? And, two, did anybody in Facebook do this kind of analysis about its own website, or were all those all external studies?

A No, those are internal studies. Like, the --

Q Oh, those were internal studies. Okay.

A Yeah. So there was a leak at some point. I have not -- I don't think I've seen this paper directly, but we could check. It might be in the cache. If you would do a Google search -- actually, let me do a super-quick one. "Neo-Nazi groups Facebook recommended."

So there's a couple different things. Oh, this is interesting. So there was a different one that got leaked in 2019, which was that Facebook allowed advertisers to target users based on topics such as "Joseph Goebbels," "Josef Mengele," like, neo-Nazi related topics.

But there are other ones that -- I can find this for you. There's a whole bunch of leaks that have happened that are related to Facebook recommending extremist groups.

Q Got it. And it sounds like Civic Integrity would've been the unit that would've done those studies for the United States?

A Yes. Yeah.

Q Okay. Great.

[redacted] Well, with that, thank you, Frances. I really appreciate it. This has been wonderful. And I'm looking forward to chatting with you again. But I'll turn it back over to [redacted] because I know he has about 10 minutes of wrap-up to do before we let you go.

Ms. Haugen. Okay. Thank you so much.

[redacted] All right. Thanks.

Let's go to document 2, [redacted] if we could.

BY [redacted]

Q And then page 2 -- well, let's start with page 1 of that document, and then we'll go to page 2. This is not a document from you.

A Uh-huh.

Q So this is a document from the Facebook Community Standards Enforcement Transparency Center. Page 1 shows that this is related to the Community Standards Enforcement report, "Violence and Incitement."

A Uh-huh.

Q And if we go to page 2, there is a description. It's 0.04 percent to 0.05 percent, which is the prevalence on violence and incitement.

A Uh-huh.

Q So what does that mean?

A Ah. Great question.

Ms. Connelly. Yoni, I'm sorry to interrupt. Are we on doc 1 or doc 2?

[redacted] Doc 2, I believe.

Ms. Connelly. I have that as the hate speech one. Is that correct?

BY [redacted]

Q Okay. Well, let's go to doc 1 then. Sorry. And this is exhibit E. If you scroll down to page 2, there's that. Okay.

A So the problem with this measure is, like, if they're using -- so, one, this is a recurring problem on Facebook. They don't release their methodologies --

Q So this is actually related to a followup point which I wanted to let you know. If you scroll to page 3, how we calculate it, it's dynamic. And when you hover over it on the website, this is the text that pops up. It's, "Prevalence is the estimated number of views that showed violating content divided by the estimated number of total content views on Facebook."

A Oh, this is great. Okay. So -- but that's the thing, though, is that word can mean different things. That's the problem.

So let's imagine two different ways you could calculate this. I can tell you which of the two that I think they're doing, but then we'll talk about why that's a bad way of doing it.

So one of the ways you could do it is, you could go and you could manually screen a whole bunch of content, like, 100,000 pieces of content. And Facebook does this kind of thing on the regular. You could go and look at 100,000 pieces of content and label each one individually. Facebook does this regularly, because they use it for a variety of different things.

A second thing you could do is, you could train an AI based on a certain number of examples of violence-inciting content, and then you could use that AI to estimate how often you saw those problems.

The problem with Facebook's systems is that they're bad at detecting things, right? Like, AI especially -- like, the things that Facebook's documents talk about, this

idea that both for hate speech and for violence-inciting content, those things require a huge amount of context that is difficult to train an AI for.

So I'll give you an example. This is from the documents. Imagine the post, "White paint colors are the worst." So anyone who's ever painted a room knows you go to the, like, hardware store and you look at, like, 150 paint chips that are all white but they're vaguely different shades of white and you know that if you pick the wrong one your room will, like, feel bad, right? White paint colors genuinely are the worst.

The problem for an AI is that it sees the words "whites are the worst," and because AIs are not smart, they're not actually intelligent, they don't understand what is and isn't hate speech.

So you end up in this situation where, if you're using an AI to estimate the amount of hate on the platform and your AI is bad at detecting hate -- the thing that I would recommend you request from Facebook is they should have to give you 1,000 samples -- so let's imagine you stack-ranked all of the scores in your system, so you have a 1 percentile, a 2 percentile, a 10 percentile, a 20 percentile, all the way up to 99th percentile score, or maybe you do it in thousandth-of-percentile for scores. If they had to give you 1,000 example pieces of content at each scoring level, you could go look through it and say, how good is this classifier? How good are you at detecting what is good and bad content?

Because I think you would see that, like, you know, right now, Facebook may consider only things that have, like, a 97th percentile score and higher as hate speech, but you might look at it and say, oh, interesting, when you look at the 90th percentile or the 75th percentile score, still, 80 percent of this content is hate speech, right? Facebook might not consider a score of 0.89 to be hate speech, but 80 percent of the content here is hate speech.

And then you could go back and say, actually, Facebook, could you recompute this score based on a cutoff at 0.65, right? And you might find that there is 10 or 100 times as much hate speech.

And I can write that out for you as, like, a little white paper. Because, like, I think that would be a fascinating exercise just to see how Facebook's own estimates of their bad content changes.

Q Okay. Thank you.

I don't think I have any other questions on this document. We're going run through some last-minute ones quickly.

So Facebook counter-arguments. I think I'm going to put aside some of the ones I was going to ask you about in particular. But some of the things that we would expect Facebook to say, because they have said it publicly in the past, is that Facebook is the best in the business at content moderation, that they do this better than anyone --

A Sure.

Q -- and, as you said yourself just recently, it's difficult, and they are the best at it, so why are we focusing this conversation today on them.

A Sure. This is a problem across the industry, right, and just because other people are doing worse at it doesn't mean that Facebook's current levels of investment are appropriate. Right?

Facebook also has the largest impact in the industry. You know, they have the most impressions, especially for this kind of content. And so it's not enough for them to be -- like, you should expect them to be doing the most. But, even then, you know, if they're making $45 billion of profit a year, right, the question here is, do they deserve to have 17-percent profit margins or do they deserve to have 15-percent profit margins, right? Because if they were investing two times as much in integrity, they would still be

radically profitable, right?

Like, after my disclosures -- so they like to say, we have invested $13 billion -- is it $13 billion or $16 billion? I think it's, like, $16 billion -- since 2016, right? And it works out to something like $2.5 billion a year. And, after my disclosures, they came out and said, we're going to up the amount we spend on integrity per year to $5 billion.

So, if you're going to make $45 billion of profit a year, you know, should you have to spend $5 billion of profit -- or $5 billion on safety, or should you have to spend $7 billion or $10 billion on safety?

And because there is no transparency, we as the public don't get to weigh in on what level of safety there is. And I think if the public had gotten a chance to weigh in and say, should Facebook make $100 million less per year, or should we have had enough people working on safety that we could've had a system where if something like the insurrection had happened anywhere in the world we would've known, I think people would unquestionably say, we should spend $20 million more per year and have that be true. Right?

Q And the decision about how much to spend on this -- if, for example, Mark Zuckerberg made what some would consider the wrong decision, is there any way for anyone to fire him?

A No, of course not. He has something like 56 percent of all the voting shares in Facebook. Like, it doesn't matter that the -- it's called the non-managerial shares. So, every single year for at least the last 5 years, the non-managerial shares have voted and overwhelmingly have voted for one share, one vote, right? Because there is no -- and they also have consistently voted to have Mark not be the chairman and the CEO. Because, right now, Mark doesn't have to talk to anyone, right? There is literally no accountability for him.

Q And he chooses the structure, so it reflects how he wants to run the organization.

A Exactly.

And the thing I'm extremely, extremely worried about is, I was told by a reporter -- so I want to be clear, this is secondhand information. Like, this reporter asked me about it, and he was like, I have heard from multiple people that Mark basically spends his days now in the metaverse. Like, the reason why he has this vision that everyone is going to spend all their time in the metaverse is he spends all of his time in the metaverse.

And I think there's a real thing of this. Mark is horribly in over his head. He has been asked to basically run, like, the mass media systems for large chunks of the world. And he keeps being told, you're causing genocides. And his solution to it is to spend his life in video games. It's a huge problem -- huge, huge problem.

[redacted] Well, on that sad note to end, I'm going to wrap, and, in that, I will allow you and your counsel to make any clarifying things on the record. But it's been a long day. I appreciate you being with us and taking the time to answer our questions.

Ms. Haugen. Thank you, guys. Always happy to help.

[redacted] Good. If something occurs to you later, please reach out to the select committee to correct the record. It sounds like we'll be having a number of further conversations, so that won't necessarily be necessary.

If you or your counsel has any questions, you want to clarify anything on the record before we recess this, I want to make sure you have the chance now.

Ms. Haugen. I'm not aware of anything.

Andrew?

Ms. Liu. I think that what we would prefer to do is to, you know, wrap and then huddle and then come back to you guys to see if there's anything that we do think needs to be clarified, and/or we can raise it in the next session.

[redacted] Sure. Okay. That's fine. I just wanted to -- I wanted to give you an opportunity now. That's fine.

Ms. Liu. Okay. After you get off the record -- I'll let Andrew talk, and then, after you get off the record, I want to say a couple things.

Go ahead.

Mr. Bakaj. Yeah. Just a quick clarifying question. So we will have an opportunity to review the transcript beforehand, correct?

[redacted] Oh, yes.

Mr. Bakaj. Okay. That's fine. So that'll be helpful as well. Okay. Thank you.

[redacted] Yeah. I was just going to say, okay, you will have the opportunity to review the transcript.

Mr. Bakaj. Okay.

[redacted] So I think, if there's nothing else we have before we recess this interview subject to the call of the chair, I think I will say it's time for us to go off the record. It's now 4:01.

Thank you, everyone.

[Whereupon, at 4:01 p.m., the interview was recessed, subject to the call of the chair.]

Certificate of Deponent/Interviewee

I have read the foregoing ____ pages, which contain the correct transcript of the answers made by me to the questions therein recorded.

_____

Witness Name

_____

Date