PrivacyCon 2020 Part 2 July 21, 2020

LERONE BANKS: OK, welcome back. So in case you are a bit drowsy from lunch, we're going to shake things up, do things a little bit differently. Our researchers today have lots of empirical data to share, so what we'll do this time is that we'll have Q&A after each presentation. So immediately following the presentation, we'll have a free Q&A with the presenter.

So for the audience, that means you'll send your questions at the end of each presentation. So please send your questions to privacycon@ftc.gov as you think of them.

With that said, let me introduce myself. My name's Lerone Banks. I'm a computer scientist in the FTC's division of Privacy and Identity Protection. Welcome to panel 4, where we'll talk about some empirical data related to specific technologies.

With that said, let's get started with Madelyn Sanfilippo from Princeton. She's presenting privacy, risk, and disaster-response apps. Welcome, Madelyn.

MADELYN SANFILIPPO: Thank you. Hopefully, we're currently on the title slide. The projects that I will discuss on behalf of my wonderful collaborators today focused on privacy issues around apps during emergency circumstances. The paper we submitted focused specifically on hurricanes and natural disasters, but I'll also discuss some implications along the way for other crisis contexts, including the current public-health emergency, as we're exploring these things in follow-up research.

Moving on to the next slide, this is pertinent, given that many factors shaping social norms and emergencies, such as individuals' inclination to share more personal information under disaster situations, as documented in many previous research studies, extends to other crises. During a hurricane or a fire, people think it's appropriate to share their location with first responders, for example, just as during a pandemic, many people are willing to share information for the purposes of contact tracing with public-health officials, although not necessarily with other actors.

As we see on the next slide, there are a variety of digital platforms to structure information flows during emergency circumstances, with government agencies both providing their own platforms and channels, as well as recommending others, in addition to the prevalent use of tools, like Facebook Safety Check. In this study, we focused on those apps that were recommended to users during hurricane season, as can be seen on the next slide.

These apps can be divided into five distinct categories-- those apps developed by government agencies, such as FEMA; those apps developed by trusted organizations that partner with public sector to provide relief, such as the Red Cross. There are also apps that are general weather apps recommended during these times. Additionally, there are hurricane-specific apps from private-sector developers that can be divided into two additional categories-- those that are transparent about their development, in contrast with those that appear, either by name or branding, to belong to government agencies, despite private development.

This latter category is problematic from a consumer-protection and deception standpoint. And many frequently change their names, though not necessarily their code or behaviors. As Apple or Google take them down from the marketplace, they simply re-enter the market with superficial changes.

Moving onto the next slide, we framed our analysis of these apps in terms of privacy as contextual integrity. This is to say we conceive of privacy as the appropriate flow of personal information in a context, in contrast with the privacy harm. That can be understood as inappropriate information flows. In this sense, information flows themselves can be deconstructed in terms of information subjects, senders, recipients, and types, as well as transmission principles, in order to compare them and understand where violations of users' expectations might occur.

We use this framework to make comparisons throughout our overall research framework, as seen on the next slide. We compare the context of privacy policies as indigenous governance and regulations as exogenous governance of information flows with actual information flows and practice, which we identified from a combination of static analysis of permissions; dynamic app analysis of flow traces, including the recipients and decryption of traffic to identify information types; as well as user experiences, described anecdotally in reviews and simulated through our own controlled experiments within virtual mobile machines.

As we can see on the next slide, the governance of disaster information flows is extremely complex due to polycentric arrangements of institutions, with different agencies having significant say, in addition to federal regulation and distinctions between Personally Identifiable Information as PII, and Sensitive Personally Identifiable Information as SPII. The key points for our purposes today are highlighted on the next slide.

Specifically, I would like to note both the ambiguities of routine uses, which is likely a source of discontinuity between government and partner organizations, whose routine uses vary significantly; and the nuance of trusted partners, including other government agencies at various levels; utility companies; hospitals; and relief organizations, from the Red Cross, to religious groups, and things like Team Rubicon. These partners are subject to restrictions, which are actually similar to those on federal agencies under the Privacy Act, including limiting redissemination and to-need-to-know circumstances. Some of you may remember that this came to be an issue around FEMA inappropriately sharing too much information about hurricane and wildfire victims with contractors in 2019.

We can move beyond governance to look at flows on the next slide. This is a visualization of information flows from the apps within our set that are sharing location information, for a variety of different reasons, with many third parties. Some of these flows violate not only regulations, such as the Red Cross sharing location of victims with Flickr and social-media companies via an installed third-party library, but many are not disclosed in their privacy policies. I will differentiate between these types of violations in a minute. But first, I'd like to briefly revisit some additional concerns raised by users on the next slide.

In addition to requirements from governance, user expectations should also theoretically be met under conditions of contextual integrity. Some users noted that user permissions, or options to control personal information, did not work on the very apps being promoted as the best to use during a hurricane. Concerns about persistent tracking were particularly significant in these complaints.

Further, others noted, in relationship to the Red Cross apps, which are depicted on the next slide, that some of the persistent tracking information was too accessible to anyone who requested it, tracking individuals in real time and indefinitely, though both of those problems have now been corrected to an extent. The concerns about the ability to track victims via Red Cross by former intimate partners in cases of domestic violence were not, however, addressed via the updates. We tested these and other complaints, as well as explore what options users actually had to control their privacy, as depicted on the next slide.

Specifically, we found that many of the apps considered provided insufficient or misleading options, with the most obvious problem being that, despite user preferences not to share location with a particular app, if they shared location with another app, it might be shared in order to personalize their disaster apps as well, in addition to personalization of other outside apps. Five hurricane apps shared location with one another, as highlighted in red. So if an individual were to share with any one of those, it was happening in all five of them. The nature of those relationships directionally is further described in our paper.

In comparing all facets of our privacy analysis, we classified the privacy compliance of all of the apps in our study as categorically represented on the next slide. Here, we differentiate between apps that are wholly compliant with policy-- those that comply either with their own privacy policy or regulation, but not both, and those that are compliant with neither. Looking more specifically on the next slide at those that are compliant, as highlighted in green, there were three apps that behaved appropriately, transmitting no personal information to any third parties, complying with all expected regulation and behaving in practice as was disclosed in the user agreement.

On the next slide, as highlighted in yellow, we see apps that did not act in ways consistent with information flows described in their own privacy policies, but that did not actually violate any laws or requirements under contractual obligations with the government. These apps simply violate user expectations.

On the next slide, as highlighted in orange, we see apps that comply with their privacy policy but that are otherwise problematic. Some of these, such as Dark Sky and Global Storms, inappropriately share data with trusted partners, though they themselves are not trusted partners. The others violate user expectations and are problematic from a deception standpoint, rather than privacy violations, as they appear to be NOAA apps, when, in fact, the National Oceanic and Atmospheric Association does not provide a consumer app.

On the next slide and highlighted in red, we see our classification of Red Cross apps as problematic due to information flows that are not disclosed in privacy policies and that violate contractual obligations with FEMA on user data. So note that when I explained the user-

experience violations in a previous slide, this is different than actually violating the policy. We have brought these to the attention of the Red Cross, and we would attribute, based on the information that we have, that these issues are primarily due to the use of third-party libraries and a lack of communication between technology and policy offices within the organization, rather than some particularly malicious intent.

Overall, the implications of this study are multifaceted, as depicted on the next slide. The study helped us to identify what aspects of context shape the unusual and under-addressed social norms that apply about information sharing. Particularly, emergencies' end and duration ought to be considered as an aspect of this context. Further, people do not normative really object to the information flows during disasters, but rather to the aggregation and reuse of this data outside of disaster circumstances or to specific actors as recipients in this context.

Finally, users ought to be able to reasonably expect what flows are going to happen in practice because privacy is contextual. However, certain events, including hurricanes and the current COVID pandemic, change expectation.

Moving onto the next slide, we build on this study to explore privacy issues around contract-tracing apps and our subsequent research, and thus far, see very similar patterns. There are, again, the same categories of apps, including government apps and privately developed government-imposter apps. There are also general health apps that have been repurposed for the current context. Further, there are also efforts by major platforms, which are sort of in parallel to the Facebook practices during hurricanes.

Yet, a really major difference is the increased objections to the potential for misuse prior to data collection or use of the app by broader sections of the public. Overall, what we see is that governance ought to apply to all parameters of information flows, not just to a small subset of actors and information types. This would provide clarity around regulatory expectations and practices so as to better inform users, and would likely meet user expectations under emergency circumstances better.

The final slide-- an overarching issue that connects this work to an emerging concern in privacy research is that of context collapse in emergency and other circumstances as introduced by digital technologies. Many of the concerns around contact tracing, as well as around Red Cross's Safe and Well program, stem from overlap of actors, information, and circumstances that people feel are inappropriate, particularly in terms of the long-term consequences.

I'm happy to address any questions that may have been sent to FTC right now, though also, feel free to reach out to me or my collaborators. Thank you.

LERONE BANKS: Thank you very much for that informative presentation, Madelyn. This is very timely research. And you actually mentioned the idea behind one of the first questions I think that we have, which is about contact-tracing apps. And so I understood you to say that, in some-- well, I guess, the first question is, how much analysis have you done with contact-tracing apps? And what has been the outcome?

MADELYN SANFILIPPO: So this is, I think, a really logical direction to take the previous study, the paper that we shared is sort of a pilot study, exploring how we could bring all of these different types of data analysis together in order to understand relative levels of compliance and violation of user expectations. What we have done thus far, with contact-tracing apps, is to collect, obviously, all of the apps and begin testing some of the user concerns that have been articulated.

However, the dynamic app analysis following traffic has not necessarily happened in every case because some of the apps we're considering haven't actually been deployed, and so we're doing preliminary analysis on some of these, based on user concerns. But we're able to categorize them according to the same parameters about who has developed them, how transparently they've developed, and what types of exogenous regulations or policies might apply to them.

LERONE BANKS: I see. And so based on some of that preliminary analysis, you are seeing at least some early signs that suggest some of the issues identified in some of the other apps that you've looked at in your paper are also starting to crop up in these contact-tracing apps. Is that right?

MADELYN SANFILIPPO: Yeah. yeah.

LERONE BANKS: Yeah. Given that, do you have some recommendations for the approach that regulators should take towards analyzing these apps? And particularly, given the fact that, unlike maybe other disasters, like hurricanes, where there is somewhat of a definitive beginning and end, the pandemic, unfortunately, does not necessarily have that clear-cut delineation. So the question is really, do you have some suggestions for how regulators should approach it? And are those suggestions different based on the nature of a pandemic, which, I think, is different from other disasters.

MADELYN SANFILIPPO: I think that's a really good point. However, I think that duration and a time element of this particular emergency context could still be addressed. So it would be my recommendation not that we think about when the pandemic context is done definitively as a whole, but rather, when individual harms that could be associated with exposure to someone with COVID actually terminate. There is an end to a period in which someone may have been infected through this. And so it's not necessarily a matter of maintaining all of that data set from beginning to end of pandemic, but rather, maintaining it only as long as is necessary in order to trace particular harms and to protect public health.

Further, I think making guarantees that this data won't be used for other purposes would be much more consistent with individuals concerns. For example, the level of trust between a public-health department and trust in particular commercial platforms is not necessarily equivalent. And so I imagine that much more compelling arguments and impetus to use some of these things could be made if the actor responsible for this data and making assurances that it will not be used for other purposes or after a period of time would be much more trustworthy from the perspective of users, at least in terms of the complaints that we're investigating, or concerns we're investigating at this point.

LERONE BANKS: I understand. Let me ask you one more question from the consumer side. So I assume that you've probably analyzed your privacy policies than a typical consumer has ever actually read, right? Do you have some advice on how consumers can read them more effectively to address those concerns-- some tips that you may have, given your comprehensive analysis, that you can advise consumers on how to find the most relevant information or interpret legal jargon?

MADELYN SANFILIPPO: Yes. It's not easy to read privacy policies at all, and that is something I spend quite a lot of time doing. In particular, one of my coauthors, [INAUDIBLE] Schneider, has done extensive research on how we might better visualize or represent this information in a way that's more easily comprehensible to users.

And so on the one hand, I would recommend that people try to communicate particular information flows that they are structuring in their policies in a standard format, as opposed to in the middle of large, complex paragraphs. That information maybe still needs to be there from a legal perspective. But from a consumer standpoint, looking at a table that says, this information is being collected this way and will be used for this purpose, is a lot more understandable.

From a user perspective right now, I think flagging particular issues that you might be concerned about or third-party advertisers, for example, and looking specifically for those things amongst the text is, perhaps, one of the most useful ways you can skim these policies without necessarily reading through all of the legal jargon yourself. You can flag particular concepts, or third parties, or uses that you're uncomfortable with and read to see if they are covered within a policy.

LERONE BANKS: That makes total sense. Hopefully, the consumers that are listening today will take some of that advice. Thank you very much for your great work.

MADELYN SANFILIPPO: Thank you.

LERONE BANKS: Next, we'll have Christin Wilson, who will present the team from Clemson's work on getting malicious skills into an Amazon Alexa skill store. Welcome, Christin.

CHRISTIN WILSON: Thank you. So good afternoon, everyone. Before I begin, I would like to thank FTC for providing me this opportunity. I would also like to thank my research team at Clemson University, especially Dr. Longsheng, Dr. [INAUDIBLE], Song, Jeffrey, and Daniel.

So we are excited to present our paper, "Dangerous Skills Got Certified-- Measuring the Trustworthiness of the Amazon Alexa Platform." So a brief introduction-- the user base of Amazon Alexa has been rising rapidly over the last couple of years, and this actually encourages third-party developers to build new skills. So here, "skill" refers to a voice app, so that's what the Amazon Alexa platform calls it.

So a skill has to be certified by the team before it's published to the end users. And a weak rating system build resulted in malicious skills entering the store. So these can be privacy-invasive, this can disseminate inappropriate information to users, et cetera. So we are especially concerned about children and the skills meant for them.

So the next slide-- we have our three research questions. Number one, we want to evaluate whether the certification system is efficient and trustworthy. Number two, do policy-violating skills exist in the skill score currently? Third, how do Google Assistant's certification system compare? Next slide.

So before we move further, let's just discuss, how can third-party skills collect data? So there are two methods. The first method is to configure permissions in the skill. So when a developer develops a skill, he can just configure some permissions.

So what happens is when a user enables the skill, a prompt will be sent to his phone-- the Amazon Alexa app on his phone-- and you have to provide permission. And this data is actually taken from the developer account, so you're not providing it. It's just taken from the account.

The second method is to collect the information through voice. So this is directly done during an interaction. So Alexa will just ask you, what is your name? And you can speak it back to them. So in this case, no prior consent or permission is taken while this skill is enabled.

So now, we come to the next slide. This is the first research question. Evaluate the certification system.

So we developed skills that violate seven privacy and 14 content-policy guidelines. So this is actually provided by the Amazon team to the developers in the developer documentation.

We do have some ethical disclosures. We have obtained approval from our university's IRB. We do not use or share any of the information collected, if any. And we remove the skill as soon as we see that it is certified. And for high-risk violations, we do try to provide a disclosure.

So the next slide discusses the first subsection. It's violation of children-specific policies. So these policies mainly focus on the collection of data from children and the content provided to children. So like in the image, you can see that you do not want a skill asking a child for personal information or encouraging him to drink, or smoke, or do something illegal without telling his or her parents.

So in the next slide, you can see that we were able to get 119 skills certified in this category. So there are two types of skills-- one that could collect data and one that would provide some inappropriate content. For the ones that could collect data, it had the following features. So it could collect personal information-- and reminder that the users are actually children, and it's completely restricted by Amazon to collect personal information from children. The second one was it could save the collected information in the developer's database. So we could save it in our DynamoDB.

The third one is we didn't provide a privacy policy. The fourth one is no prior consent is taken from parent or guardian before the collection of data. So when the skill is enabled, nothing is told to the parents, and when the child uses it personal information will be taken from them.

And then no access is provided to view, delete, or modify the collected information from our data set. And also, the developer-account details that we used were fake, so they can't actually contact us to ask about the collected information. The other skills that were published had content not suitable for children or encouraged them to use services outside of Alexa.

So the next slide contains the next two subsections. These are a violation of general content guidelines and privacy requirements. So these policies are mainly for the general audience. And we were able to get 115 skills certified.

So these either had the data-collection violations, just like [INAUDIBLE] children skills. Or it had some data that is not supposed to be told, like there were, like, health-related information, promotions, disturbing content, advertisements, promotion of alcohol, drugs, illegal activities, et cetera.

So in the next line you can actually see an example of a skill that we submitted. You can see that the skill [INAUDIBLE] stories was live on the-- it ask for the user's name. It got the full name, and generated a story with their name. But you can see that no privacy policy was provided. It's a kids-category skill, and you can see that the name was actually saved in the DynamoDB database. So we made this just for illustration purposes, and we ourselves provided that name, and it's not actual user data.

And next slide-- so the experiment results-- we've been able to certify 234 skills in total. So it's not 234 unique skills, but it's like 234 different certifications, I would say.

So it was conducted over about a year. And we did have to resubmit some of these skills. So initially, some of these skills were actually rejected, so what did we do? We just had to use a simple counter to delay the session in which our privacy-policy violating response is [INAUDIBLE].

So if we set the counter as 4, the first four responses from [INAUDIBLE] will be perfectly fine, and the team would actually certify the skill based on that. And after that, since the counter was four starting from the fifth, the policy-violating response will be [INAUDIBLE].

So in the next slide, we are actually discussing our observations. So the first observation is the inconsistency in checking because we got different responses for the same exact skill each time we submitted it. The second one is limited voice checking. So they're not to looking at the code or anything. They're just talking and just seeing if the conversation matches.

The third one is over-trust based on developers. So this is actually [INAUDIBLE]. We marked that the skill does not contain advertising, but the skill actually contains-- so the certification team, instead of actually checking for it, they just trust the developer and certified the skill. The fourth one is humans are involved in certification, and it's not an automatic process yet.

The fifth one is negligence during certification. So this is because, for some skill sets, especially the ones that can ask for the name from users-- this was actually a story skill, and this was asking for the name in the first session itself. But we got some rejections for that. The certification team

actually had a probably what the content of the story. The story had some violence in it or something. But they never complained about us asking for the personal information without providing a privacy policy, and that, too, from kids. So this shows a negligence from their part, I would say.

Next slide. So second research question was to look for existing policy-violating skills in the store. We only tested 825 skills. There are about 100,000 skills, which we can't actually check. So we just get skills that either had a negative review or had a privacy policy provided.

So by looking for skills that had a privacy policy provided, what we wanted to do was, like, Alexa only requires skills that collect personal information to include a privacy policy. So this was our assumption that they might be collecting personal information-- made us look through them. And we identified 52 skills with possible privacy violations. Again, we use the word "possible" because we can't really ensure whether some policy violations actually existed, because we can't access the code.

There was also 51 broken schools that didn't work. So there is no constant check being done to see if the skills are looking perfect.

So the next slide-- we have a few examples about some privacy-policy problems we saw during the manual testing. So the image on the right is actually a skill developed by Amazon, and it's actually a weather app, and it's available by default on all Alexa devices.

So it mentioned in the description that it collects the user's device location, like any other weather app would do, but it does not provide a privacy policy in the usually allotted space. So since this is an Amazon-developed skill, it's OK because you can actually find one in the bottom of the page. But there are other skills that are not developed by Amazon and mention about collection of data in their description, but don't really provide a privacy policy.

The other [INAUDIBLE] an example of a badly written privacy policy. We don't really know what the developer actually meant by that line. There are also examples of privacy policy URLs leading to the Google search web page, other developers' privacy policy, et cetera. So these were links provided before certification. So during certification, the team could actually see the URL, and they just might not have gone through it or just neglected it.

Next slide. We also did a preliminary comparative measurement on the Google Assistant platform. So we got 15 out of the 85 [INAUDIBLE] certified. And for general actions, we got 101 out of 185 certified. So actions is the Google equivalant of skills.

This data just suggests that Google's vetting is better, but again, this is a preliminary study, so we can't really state that. We did see some inconsistency in feedback here too. And the post-certification vulnerability exists here as well. So this vulnerability means that once a skill is certified, you can make changes, and then it will be deployed to the live audience without requiring a recertification. So yeah, this still exists in both Amazon and Google, and I think this has been discussed in some other papers as well.

We did manual testing on the 76 kids sections as well-- or they call it "actions for families"-- and we found one problematic action. So this goes to say that there are policy-violating still actions in the Google directory as well.

On the next slide, we have a responsible disclosure. We have reported our findings to both Amazon and Google. The Amazon security team is still working with us on investigation and resolving this issue, but it's still going on. It's not resolved yet. The Google team, on the other hand-- the counter-abuse systems are actually issued [INAUDIBLE] part of the Vulnerability Reward Program for our work.

And coming onto our final slide, you can see that we have provided a website link. More details and video demos and actually provided in the website. You guys can take a look at that. If you have any questions, I can take them now.

LERONE BANKS: Great thank you very much, Christin. So first, for the audience, if you do have any questions for Christin, please do email us at privacycon@ftc.gov right now if you have some questions. He presented a lot of information, and hopefully, you have lots of good questions.

So while we're waiting for a few audience questions to come in, I'll ask you a question that kind of starts at the end of your presentation. You mentioned that you reported these results to Amazon and Google. First, congratulations on the award from Google. That's an accomplishment in and of itself. But sure, my question is about Amazon's response.

So how receptive were they initially, I guess, to your findings? And what was their feedback-- and particularly, your claim about the ease with which it is to circumvent their process? And how did they address the issues that you raised?

CHRISTIN WILSON: So I would say they were very good in our results. They got into a call with us as soon as they got the email. We did have a lengthy meeting discussing about what work we actually did, how we it was. Even they asked, like, did you guys try some other method? And we were like, no, this was collectively very easy to do this, so we didn't have to go for harder techniques and stuff.

But yeah, they're were really eager to know about our work. They are still investigating. They have tried to get as much information from us, so they're actually looking at the certification log of every skill that we actually published. So I think they are doing a good job, but it's still being done, so we don't know the final result yet.

LERONE BANKS: I see. Well, at least it sounds like it's a pretty collaborative process, which is not always the case in these types of instances. So I think it's good to see that.

Let's see. I think there might be a question coming in, but another question is about static analysis and whether or not static analysis would have been effective in identifying your techniques for bypassing their checks because, I think, in your paper you mentioned that Amazon didn't-- or would the architecture prevent static analysis of certified skills, or that they don't do

static analysis. So would should Amazon do static analysis? And would that have been helpful to you?

CHRISTIN WILSON: So with the current architecture, what actually happens is the back end is completely invisible to the certification team, I would say. Because this is a black box, we still don't have confirmed results. This was a question that we raised to Amazon, and they haven't responded to that yet.

So from what we know, they do not have access to the code-- the back-end code of the developer. So they have no means to actually check it right now. So all they do is, actually, they talk to the skill, and they look at the responses that are coming in. And then they decide.

So yeah, this was one of the solutions. You should be taking permissions from the developer to actually view the back-end code and maybe just block it from being made. The developers can change it any time right now, and it will be deployed to the live audience. So this is not something that should be done. It should be blocked is what we are saying.

Many other researchers also said this in other papers-- that developers should not be allowed to change the code. Until recently, the developers could only change the back-end code. The front-end code could not be changed, but recently, they changed that too. So now, you can update both the front-end code and back-end code without requiring a recertification.

LERONE BANKS: OK. I want to make that point clear. So it sounds like what you're saying is that there's a pretty significant blind spot that Amazon has-- or third-party code-- and that the certifiers within Amazon cannot actually see the code. And if the developers make modifications to that code, that does not have to get recertified. Can you really make clear the significance of that blind spot, in terms of what the potential vulnerabilities that can arise from that, are?

CHRISTIN WILSON: So with the blind spot, what the problem is-- even if you create a chat bot that can go test the skill for 1,000 times, the malicious thing can happen on the 1,001st session. So you can't actually find it. So unless you have the back-end code, you can't actually find out all the policy violations. And again, because of the skill certification vulnerability, even if the certification system is really good in detecting all these problems and rejecting all the skills, a skill can actually pose as a good skill initially, get it certified, change the back-end code, completely change its functionality, and then just-- yeah, it doesn't make sense.

LERONE BANKS: I see. So that sounds like maybe an opportunity for regulators to step in and say that if you're going to offer skills or that any organization will offer skills, then they should be able to have access to the code in order to do a comprehensive analysis before it's made public.

CHRISTIN WILSON: Yes. Again, like I said, this is still an assumption because that's what we know-- that it's a black box. We have contacted Amazon, and they haven't responded to us about this yet. But from what we saw in our experiments, we can say that they're not looking at the code. Otherwise, they would have definitely found some of these mistakes.

We usually use to name their variables as first name, last name. And if they just look at the code once, they can see we are collecting the full name of the user. So if they actually looked at the code, we would say that they would have definitely found this. But yeah, still, this is an assumption.

LERONE BANKS: OK, I understand. Let's see. We have a question from the audience. Can you elaborate more on what you found with the skills and the kids category?

The question asks whether or not there were specific violations of COPPA. I don't know if you're familiar with the details of COPPA. But even if you're not, can you talk about what you did see, specifically, in terms of what type of information was being collected within kids skills and what you perceive the violations might have been?

CHRISTIN WILSON: So I think, regarding the collection of data, it was mostly collecting either the device location, or the user's location, or the user's name. So I did see a lot of stories skills asking for a name to personalize the stories. It just makes it interesting, I would say, to get a story based on you as the character.

So that was mostly it. But again, we tried providing the full name and the skill [INAUDIBLE] with it. But we do not really know if the skin was actually collecting the full name because, I think, personal information is going to be the full name, according to COPPA, and not just the first name. So like I said, since we don't have access to the code, we don't really know what they are collecting. Are they keeping both the first name-- or are they just taking the first name all those kind of things?

LERONE BANKS: I understand. So it sounds like that's an area for some closer analysis.

CHRISTIN WILSON: Yes.

LERONE BANKS: So thank you very much, Christin. For your work.

Next, we'll have Aerin Zhang. She's here to present CMU's research into consumer attitudes about video surveillance and facial recognition. Welcome, Aerin.

AERIN ZHANG: Thank you, Lerone. I appreciate this opportunity to present our research at PrivacyCon. So today, I'll present our work on understanding people's privacy attitudes towards video-analytics technologies. This work is part of the Personalized Privacy Assistant project.

There were 17 million surveillance cameras in the US in 2018, and if that number is not impressive enough, 1 billion cameras are expected to be deployed globally by the year 2021. The massive amount of video data captured by these cameras motivates video-analytics technologies, which use computer software to automatically process and understand videos.

Such technologies have been greatly improved due to recent events in deep learning and computer vision, and they have become increasingly sophisticated. Such software can be easily applied to real-time IP cameras or store footage from any cameras. Those analyses often happen

without subjects' awareness or consent. Important information about the data collection-- like how long the footage is retained, whether the information could be shared with other entities, or the purpose of analysis-- is often not available to data subjects.

Piracy regulations, like GDPR, include stricter laws to govern the use of video analytics. The regulations require entities that use video analytics notify data subjects and enable them to opt in or out of some practices at or before the point of collection.

But there are several different types of video-analytics technologies today. Facial recognition is the most prominent type, and also, has several variations. It can identify an individual by matching an image of a person to a database of known people. There's also anonymous face detection that can be used to estimate demographics of the person.

Another type is facial-expression recognition that detects individuals emotions. Other than facial recognition, [INAUDIBLE] detection is also one type of video analytics. This image shows how the software is analyzing the video feed to count the number of passengers in the subway compartment. Next slide, please.

The gap between the current disclosure practices and the requirements of the regulation draw our attention to the lack of guidance on how to do a better job at communicating these data practices and what choices to expose to data subjects. In order to facilitate appropriate notice and choice about these different types of data-analytics deployments, we first want to understand people's privacy expectations and preferences with regard to these deployments.

We ask the following research question. Do people know about these deployments? And how do people feel about them? Especially, we are interested in people's surprise levels, whether to expect these practices at certain places or not, their comfort level, their notification preference-- meaning whether they would like to be notified or not, and if yes, how often they want to be notified. We're also interested in whether people would allow or deny those practices if given a choice.

With these research questions in mind, we designed an experience-sampling study. So the experience-sampling method is a longitudinal research methodology which enables us to engage and survey participants in the moment as they go about their normal daily lives. As a result, this method allows us to collect higher quality, more ecologically valid research data than static online surveys.

In total, we collected detailed responses for more than 2,300 deployment scenarios from 123 participants. When recruiting, we tried to avoid convenient sampling of students and try to reach out to the local community. We ended up with a rather diverse sample. Here's a pie chart showing the different occupations of our participants.

This study is approved by Carnegie Mellon's Internal Review board and by the following agency's Human Subject Protection office. Next slide, please.

So we first did an extensive survey of news articles about real-world deployments of video-analytics technologies. We identified four major categories in a variety of contexts. The first category is for security, which includes automatically detecting petty-crime scenes-- like pickpocketing, breakins-- or using facial recognition to identify known criminals and bad actors.

The second important type is for commercial uses. It's been used to count the number of people in a facility in order to optimize operation, like staff management. Or it's used for targeted advertising based on demographics, individual profiles, or reactions when people are looking at items. Yes, you can be advertised based on what you look at and your facial expression. It has also been used to read people's engagement in museums, movie theaters, and comedy clubs.

The third key usage revolves around identification and authentication. Facial recognition can be used to replace work IDs, membership, and loyalty cards. It has been used to track attendance at gyms, schools, workplaces, and even churches.

And the last category of uses is more advanced. Facial recognition and emotion analysis can be used by health-insurance providers or hospitals and doctors to make health-related predictions, and by employers to evaluate employees' performance and monitor their productivity. In total, we identified 15 unique purposes and a baseline purpose, which involves only generic surveillance with no video analytics.

Here, I'm going to briefly explain the study protocol. The participant participants download and install the study app on their own Android devices, and the app first asks them to fill in a pre-study survey. After that, participants were instructed to go about their regular daily activities, and the app sent them push notifications prompting them to complete a short survey based on their current locations. Because the GPS location is not precise enough indoors, participants first confirmed the place they were at by selecting from a dropdown list of nearby places. Then the app displays a video-analytics deployment scenario relevant to the place they were visiting, and then they answer four in-feature questions about their surprise level, comfort level, and notification preference with regard to this scenario.

On the days participants receive push notifications through the app, they also got an email in the evening to complete a daily summary. The summary asked participants to revisit notifications they received during the day and to provide additional responses. The process will happen for 10 days, and participants finish the study with a post-study survey.

Now, I'm moving on to some of the results we found in our study. Due to the length of the presentation today, I'm only showing some of the results, and more can be found in our paper.

So this slide shows a summary of participants' comfort levels organized around 16 different purposes we previously identified. It is clear to us there is no scenario where everybody feels uniformly about. People's responses vary greatly for each purpose.

For instance, scenarios related to security appear to surprise participants the least. Close to 72% would feel somewhat or very comfortable about these scenarios.

On the other end of the spectrum, we observed considerably less acceptance by event scenarios, like health and productivity predictions, where only 70% feel somewhat or very uncomfortable. Participants are least comfortable with employees making predictions about their work productivity.

So after the 10-day study, 75 out of 123 participants grew more concerned about these practices. 80% of these 75 participants developed stronger awareness of the possible deployment of video-analytics technologies as they received notifications on their phone every day. They were not aware that video analytics could be used for so many purposes at such a diverse set of venues and with this level of sophistication. When participants commented, some of the scenarios and brought up the technology you mentioned I had never considered freaked me out,

27% emphasized the privacy issues of these technologies, like the lack of notice or consent. 25% expressed concerns about specific usage of these technologies. One said, I didn't realize I could be marketed to based on what I'm looking at in a store. I found this whole practice disconcerting. 4% were worried about implications like how the data is shared, what could be inferred from the data, and the potential abuse. Next slide, please.

So just to give a complete picture, I'm going to show some opinions of participants who stayed equally concerned or actually grew less concerned over the course of the study, even though they are minorities. So 27% of them claim they are already familiar with these technologies. 23 were not bothered by the practices. They said something like, if you're not a criminal, you shouldn't be worried about facial recognition.

And 21% expressed some level of resignation, describing the technology as ubiquitous and out of their control. 15% did not believe that the scenarios showed to them were real. And 13% who learned the benefits of these technologies become more accepting.

Now we move on to the next slide, showing results on those adhesion preferences. We asked participants, how would you want to be notified? The choices range from notifying you every time to, do not notify me. Again, we observed that people show diverse notification preferences.

This graph shows how their preferences changed before and after the study. More than half ended up with different preferences, and the majority are looking for some type of selective-notification solution instead of being notified every time. Next slide, please.

So interestingly, we observed that people grew more concerned in general, but opted for less frequent notifications as time passes. This change in preferences is attributed to some level of privacy fatigue as people got a better appreciation of the number of times they are likely to be notified. So one participant described their fear for privacy fatigue, as they received many notifications. Next slide, please.

Even with our 10-day study, we already observed privacy fatigue. So remember the regulations which expect people to manually opt in or out of video analytics each time to encounter such functionality. But because of the increasingly widespread deployment of those softwares, this could result in an unrealistically high number of privacy decisions.

The natural question to ask is, how could we reduce user burden and assist users in making privacy decisions? So I want to first provide some context of how obtaining consent works with video-analytics data collections. But there are some recent technical advances that made it possible to obfuscate people's faces in real time, allowing people to opt out of video analytics.

There are also academic efforts to build a privacy infrastructure and a privacy-assitant app for internet of things. Such an app running on people's smartphones would alert users of nearby IOT sensors, for example, cameras with video-analytics software enabled, and present them with potential choices, like opt in or opt out. However, with all the efforts, the high user burden remains a problem.

So with the data collected from our study, we're able to use clustering techniques to reduce user burden. We first group likeminded users to generate privacy profiles and then leverage clustered profiles to make predictions of people's allowed or denied decisions.

So using this method, we're able to predict 94% of the allow/deny decisions with 89% accuracy. Next slide, please.

So it is worth taking a closer look at the clusters of the likeminded subjects identified by our clustering algorithms. This graph shows privacy profiles of six clusters. Each cell represents whether people in this cluster allow or deny data practices for a specific purpose. The color blue means "allow," and the red means "deny."

Darker colors indicate a stronger cluster consensus. We see a cluster 1 and 5 of polar extremes. Cluster 1 are privacy conservatives, while cluster 5 are mostly unconcerned. They merely allowed data collection for all purposes.

The other four profiles are more nuanced, with a mix of red and blue cells. For example, cluster 4 mostly allowed data collections for security purposes and denied others.

So I'm going to summarize what we have learned. Through the study, we observed people's lack of awareness of video analytics and their desire for greater transparency. The current practices of notifying people by putting signage that states "This area under surveillance" is not sufficient and also, not compliant with regulations. People want to know when they are subject to video-analytics technologies.

We also saw how participants can be overwhelmed by the number of privacy decisions they might encounter as required by regulations, and they are looking for selective-notification solutions. Lastly, we demonstrated the feasibility of reducing user burden through machine learning by predicting the majority of decisions with high accuracy. And this concludes the presentation.

LERONE BANKS: Thank you very much, Aerin. Again, for the audience, if you have any questions for Aerin about what consumers have to say about video surveillance, please send your questions to privacycon@ftc.gov right now. We have a few minutes left.

All right, so thank you again, Aerin, for that work, and thank you for spending time talking to consumers. I'd like to spend more time doing that too.

Let me ask you one question as we wait for some audience questions to come in. In other contexts and other privacy research, we often hear the term "privacy paradox" thrown around, and other people use it too. In your interactions with consumers, did you observe any privacy paradoxes or any counterintuitive behavior responses?

AERIN ZHANG: I think the privacy fatigue that we described had something to do with the privacy paradox, but the privacy paradox deals with actual behaviors that we, in the study, did not really measure. So we are basically asking their opinions. So the privacy paradox describes the discrepancies between the actual behaviors and their saying that they care about privacy.

LERONE BANKS: I see. I have one other question for you. Oh, in your study, comfort is strongly correlated with allow and deny decisions from consumers. Is that right?

AERIN ZHANG: Yes.

LERONE BANKS: And were you able, based on consumer responses, to get an understanding of what things companies could do that were the most effective at increasing their consumer's comfort level? Essentially, what I'm asking is, are there things that companies can do to increase consumer comfort and reduce surprise?

AERIN ZHANG: I think by listing a lot of the attributes, like the purpose for which this data is collected and for how long the data is retained-- by disclosing those informations companies will receive more acceptance from users because we have seen that people, once they know the benefits and they know the whole picture of how facial recognition is used, they become more accepting to some level.

AERIN ZHANG: And presumably, that would be outside of the privacy policy that they're less likely to read, and maybe somewhere prominent and easy for consumers to understand, hopefully, right?

AERIN ZHANG: Yes.

LERONE BANKS: So thank you again, Aerin. And I'd like to thank all of our researchers today. You're doing great work, and it's informed me a lot today. And I hope our audience got as much out of it as I did.

So thank you very much, and I think we have another panel coming in immediately after us. Thank you again, and thank you to the audience for your attention.

DANIEL WOOD: OK, welcome to the fifth panel of this year's PrivacyCon. The topic of this panel is international privacy. I'm Dan Wood. I'm an economist in the Bureau of Economics at the Federal Trade Commission and the Division of Consumer Protection.

With me are four panelists. The first is Guy Aridor. He's an economics PhD candidate at Columbia University. And the research he's going to be talking about today is about how the European Union's General Data Protection Regulation, or GDPR-- how its opt-in requirement affected the mix of consumer data observed by intermediary web services.

The second panelist is Garrett Johnson. He's an Assistant Professor of Marketing at Questrom from School of Business at Boston University, and he's going to be talking about how GDPR affected market concentration among web-technology vendors.

Our third panelist is Jeff Prince. Jeff is a professor of business economics and public policy at the Kelley School of Business at Indiana University. He's also the Harold A Poling chair of strategic management and co-director of the Institute for Business Analytics at the Business School. And his research he'll be presenting is about measuring individual's valuation of online privacy across countries and also, across privacy domains.

Christine Utz is our last panelist. And she's a PhD student at the Chair for Systems Security at Ruhr University Bochum in Germany. And her research is about how design choices in GDPR consent notices affect how users interact with those notices.

So without further ado, I'll turn it over to our first presenter, Guy Aridor.

GUY ARIDOR: Yeah. Thanks, Daniel. So today, I'm going to talk about the effect of data-privacy regulation on the data industry. This is joint work with Yeon-Koo Che at Columbia and Tobias Salz at MIT.

So the fundamental tension at the heart of data-privacy regulations is that on the one hand, consumers increasingly want control over the data that firms collect on. It's just been amplified in recent years as a result of a number of high-profile data breaches, as well as an increase in the scale and scope of data that firms collect on consumers, rendering consumers to be unable to understand what kinds of data is collected on. On the other hand, firms are becoming increasingly reliant on this consumer-generated data. There's a worry that such privacy regulation might impact their functionally.

There's two main uses of data in the digital economy. The first is that this data is the fuel behind a lot of the machine-learning technologies which are becoming more and more deployed in the digital economy. And second, they're crucial for the targeted advertising-- just how many websites derive their revenues.

So next slide-- what did we do in this paper? So we look at the European Union's General Data Protection Regulation. In particular, we focus on the consent aspect of the legislation, which gives consumers additional control over the data that firms collect on. And what we try to answer in this paper is try to empirically say something about the tension I previously discussed.

First, we try to ask, do consumers make use of the privacy [INAUDIBLE] provided by GDPR? And then how does this impact the overall pool of data that firms observe, and how does this

materially impact the firm's ability to predict consumer behavior and accrue advertising revenues? Next slide.

So the empirical setting for this paper is the data is provided to us from a third-party intermediary in the online travel industry, which spans the majority of this industry across the globe. Studying this intermediary's ideal to setting the consequences of the GDPR and data-relaint firms for several reasons. The first is that the data that this firm collects is directly at the heart of the consent portion of GDPR, such that properly implemented consent should allow consumers to opt out of data collection from this intermediary.

Second, the primary businesses of this intermediary is to collect user search and purchase histories and to predict whether or not consumers are going to purchase a flight or hotel, and then conditional on this prediction-- show some advertising, which is how most of their revenues come about. And so in particular, we observed the following.

We observe a high degree of consumer search histories, we observe advertising revenues, and we observe the output of the proprietary machine-learning outcome, which are all the necessary outcomes to talk about the original tension. Next slide.

So what do we do in this paper, in terms of our our empirical strategies? So we'll use a relatively standard tool from economics, known as difference-in-differences, which allows us to get at the causal impact of the policy. And in particular, our treatment group here are the travel agencies in major European countries, and the control group here are non-EU countries. And our analysis revolves around the GDPR implementation date, which was Friday, May 25, 2018.

And it's important to point out that our specification will allow us to look at the causal impact of the policy overall, and not necessarily on particular manifestations of the policy, such as what we discussed later in this path. And so we look at the period from beginning of April, 2018 until the end of July, 2018. And I'm going to report to specifications here. One is just going to give the overall causal effect over this time period, and the second is going to give a time [INAUDIBLE]. OK, so next slide.

The first thing we do is try to use our specification to indirectly measure consumer opt-out. So it's important to understand how GDPR opt-out manifests itself in the data that we see from the intermediary. In particular, when a consumer opts out of data collection, their data is not showing up in the database at all. And so what we do is we measure opt-out indirectly by estimating the difference between the observed users and the number of users that would have been observed, had GDPR not been around.

And so to measure this, we're going to consider the following outcome variables. One is the total number of cookies, and the second is the total number of recorded searches. And so again, this will allow us to indirectly have an estimate for how many consumers opt out, but it's also going to tell us about how the overall scale of the data that the firm sees changes. Next slide.

So this is the time-variant specification. So you can see a sharp drop at week 22, which is exactly the week of GDPR implementation. And you see a steady decline of roughly 10% to 12%, and this is consistent across the different outcome variables. Next slide.

OK, so now what do we turn to? So we've established that there is a 10% to 12% drop in the total number of users that firms observe.

Now, our next question was, are there any changes in the composition of the users? And to do this, we look at a measure of persistence. So what we do is fix a website, J, and fix a week, T. And then we collect all the cookies that that website observes in time T. And then we ask, what fraction of these are still around one week later, two weeks later, three weeks later?

And we ask, what happens after GDPR? Are the resulting consumers more persistently identifiable? Next slide.

So what we find is, again, you see a sharp increase at the onset of GDPR. And the effect size of this is roughly around an 8% increase in persistence. So I'm not going to go into the details here, but in the paper, we were curious, what's the mechanism behind this increase in consumer persistence? So next slide.

So our main conclusion is really that it's important to distinguish between different means of privacy. So before GDPR, consumers could do things such as delete their cookies or use private browsing. And what would happen here is if the consumers data would end up in the firm's database, but with a new identifier, whereas under GDPR, such data is eliminated completely. And so a substitution between these may lead to a different data-generating process and longer consumer-search histories. Next slide.

So this figure illustrates exactly what I'm talking about. So in the far left, you can see the identifier column is the identifier that the intermediary is observing for a particular user. And then in the other three columns, you're going to see consumer histories. So if you focus on the first panel, the full-visibility panel, that gives the true data, that if the firm perfectly observed everything is what they would see. And so what you see is four distinct consumers with distinct search and purchase histories.

In the middle panel is the obfuscation regime, which is the pre-GDPR. And so let's suppose the first three consumers-- they don't change their behaviors at all. But suppose the fourth is privacy conscious, and so periodically deletes cookies or uses [INAUDIBLE]. So now what happens with this guy is his identifier is now partitioned into two users. So the intermediary thinks that it's saved two people, but it's actually one person. And as you can see, consumer 4 and consumer 1 now have identical histories, and consumer 5 consumer 2 have identical histories.

Now, what happens under GDPR? So this privacy-conscious consumer can now opt out of the data. So you see that the firm only observes three users now, but they arguably have cleaner identifiers. So there's two takeaways here. One is moving from the second panel to the third panel is going to mechanically increase persistence. And the second is that it might actually help

the firm predict consumer behavior because they have clear user histories, and we have an extended discussion of this in the paper. Next slide.

So now what we want to do is we want to look at what happens to advertising revenues? And so all we're going to report here is just the results of our specifications without many more details, but it's important to contextualize the advertising setting here. So we're not thinking about behaviorally targeted advertising where advertisers are bidding directly on consumer histories. We're thinking here of keyword-search advertising-- so similar to Google-sponsored search.

So an example is advertisers are bidding on consumers who search from a flight from New York City to LA. So any changes to bidder behavior are reflecting the average value of a consumer. Next slide.

OK, so these are results using the same specification as before. So first, what we find is the total number of advertisements that get clicked on has a similar effect-size drop as we saw before, which is roughly 13%. Next slide.

So we look at revenue. And so revenue is a bit interesting. So I don't report the time-varying graph here, but what we see is there is a sharp decline at the onset of GDPR and a slight increase afterwards.

So we find a negative-point estimate, but it's relatively precise and statistically significant. And the reason why, we think, is because-- if we go to the next slide-- the average bid for a consumer [INAUDIBLE] GDPR actually increases, which points to the fact that advertisers had a higher average value of consumers after GDPR. And so this partially offsets the loss from opt-out but not completely. OK, so that was in advertising.

And then, finally, we're going to turn to prediction. So yeah, so they should be on the consequences for prediction slide. So we asked, what's the impact on predictability of consumer behavior?

So I think this is interesting for two reasons. The first is obviously directly impacts from revenue in terms of the amount of personalization and product quality that they're able to offer. But I think the second reason why one should be interested in consumer behavior-- or in predictability-- is also from a consumer-privacy standpoint because it's privacy in the modern age, at least colloquially, is not just about, what kinds of information do people have about me, but what can firms predict about my behavior? And so we think that this exercise is interesting to look at from that light as well.

And so the key idea here for interpreting our results is that the privacy decisions of others affect my predictability right. So all our data gets pulled together. So if the firm has less data from opt-out, this is going to impact the firm's ability to predict for a consumer that opted in.

But the second thing is, if you think about our results from persistence and the stylized diagram I showed you before, it might actually be possible that the substitution from cookie obfuscation to opt-out actually leads to cleaner identifiers than exerts an externality to our consumers by

making them more predictable. And so in our setting, we can use the same special specification as before precisely because the firm trains their prediction model for each site on whatever data they approve from that site, which means that changes in data from one website don't affect the firm's ability to predict on another website. And so what do we find? Next slide.

So we do a short-run exercise, which we just put it through the same specification as before. And we find that there is slight improvement in prediction, but the big takeaway, I think, for us-- this prediction didn't get substantially worse, according to the measure utilized by the intermediary.

Now, we were a bit worried that the short-run effects might not give enough time for the intermediary to adjust its prediction rhythm, and so what we do is we do a back-of-the envelope logarithmic exercise, where we take the changes we saw from our earlier difference-in-difference estimate in the change in the overall scale and longer consumer search histories. And we asked, how should those affect prediction? And what we find is a roughly, a similar result. OK, next slide.

Well, what did we do today? So we looked at the impact of GDPR on a number of different outcome variables. I think there's two high-level takeaways that are very closely related. The first is I think we highlight how government-mandated privacy protections do interact with other privacy needs? And this can be important for understanding the value of such regulation. And second is that we highlight that the consumer-privacy decisions have externalities on other consumers, which is not something that legislation such as the GDPR group really think about.

And finally, just in terms of welfare, going back to the attention you talked about before-- do consumers benefit? Privacy-conscious consumers clearly do. For the others, it depends on the alignment of the preferences of firms and consumers.

Do firms suffer? Firms lose a significant number of consumers from opt-out, but remaining consumers are higher-value, so it's not wholly negative. And finally, the ability to predict is not substantially worse. OK, thanks.

DANIEL WOOD: OK, great. Thank you, Guy. Next up is Garrett Johnson.

GARRETT JOHNSON: Thanks, Dan. I'm honored to be back this year to represent our second GDPR paper with the same set of co-authors. It's with Scott Shriver at Boulder and Sam Goldberg at Northwestern. Next slide, please.

Our main research question is, can privacy policy hurt competition? Now, there's a theoretical tension between privacy and competition policy, but this claim lacks empirical evidence. One reason for this tension is economies of scale, that larger firms may have more resources to comply with regulation.

We propose a novel mechanism, though, which is B2b choice of data vendors. That is, if privacy regulation pushes firms to limit data sharing, firms may prefer to keep their larger vendors because these ventures have better products. Next slide.

So as in the last talk, we're setting the GDPR, which is a landmark privacy policy and a leading example to the world. And we, too, are going to use its enforcement deadline of May 25, 2018, as an event study.

Now, the GDPR is very complex, but its many elements contribute to increasing both the logistical cost and legal risk associated with processing personal data. And this is going to have important consequences for the web. We study the technology-vendor industry that provides an ecosystem for the web to thrive. Specifically, these vendors help websites to monetize themselves with ads, to load and share content, as well as measure and optimize site traffic.

Now, in order to provide many of these services, vendors often have to share what the GDPR considers to be personal data. And as a result of this, the industry has faced intense regulatory scrutiny with at least three EU countries releasing major reports or statements criticizing the industry. But so far, the regulators have not issued any fines. Next slide.

So today, I'm going to briefly discuss our data and then discuss our results in three stages, talk about the GDPR's impact on vendors, its impact on concentration, and then differences by website. Next slide, please. So we begin with our data. Next slide.

When you visit your favorite website, your browser interacts with the first prior domain at that site. So in this example, I visited theguardian.com, my browser's interacting with that domain. Next slide.

At the same time, your browser is interacting with potentially, dozens of third-party domains owned by vendors, selected by the website to provide these services. So here, I've used an extension for Chrome called "Disconnect" that allows me to visualize these vendors. And you can see many familiar logos, including Facebook, Twitter, and Yahoo. Many of these vendors are helping to monetize The Guardian with ads.

Now the GDP is challenging the study because normally, we cannot observe how firms use and share personal data. However, in this instance, the function is being outsourced to the browser, so we are able to observe a website's network of vendors. Next slide, please.

Our data collection precedes as follows. First, we use a VPN service to simulate ourselves as originating from within the EU, specifically from France. Second, we use a specialized piece of software, called Web X-Ray, developed by a researcher at CMU named Tim Libert. And that allows us to record all the third-party domain interactions when we visit a website.

And finally, we repeat this for 28,000 top sites regularly throughout 2018, and these sites in our data sample are the top 2,000 websites in each of the 28 EU countries, as well as the US, Canada, and globally. Next slide.

So for our results, we begin by looking at the GDPR's effect on vendor use. Next slide.

So this figure shows the average number of vendors per site over 2018. And immediately prior to the GDPR, sites use 14.4 vendors on average. One week later this falls to 12.4 vendors, which is

its lowest level. And this is a 15% reduction in vendor use, which we referred to as a short-run effect of the GDPR.

Now, obviously, would have preferred to collect a longer [INAUDIBLE] period, but we know from auxiliary data and related research that the pre-trend here is flat. Furthermore, websites appear to have waited to the last minute to make changes to their website, which is why 3/4 of the drop in vendors happens within just a few days of the enforcement deadline.

The reduction in vendor use is short-lived, , however and erodes by the end of 2018. The post-GDPR growth may just arise from a dynamic market that's expanding over time, but I'm going to show you some evidence later that this growth is consistent with sites' beliefs about enforcement falling over time in the absence of fines. Next slide.

So now, we've seen that vendor use falls post-GDPR. We're now going to turn our attention to concentration. Next slide.

The fixed ideas-- we know that vendor use falls post-GDPR. And most vendors are actually worse off post-GDPR, in terms of the number of sites that they're working with. But here, we're, instead, asking a different question, which is, do the larger vendors get a larger share of the smaller pie after the GDPR? Next slide.

Now, in order to measure market concentration, we begin by defining market shares, and our market-share definition relies on reach, which is just the number of web sites that use a vendor. So in the sidebar example, you can see that Google Analytics has a reach of two sites, and Adobe Analytics has a reach of one site. And then to calculate relative-market shares, we just take the vendor's reach divided by the total reach so that in the sidebar example, Google Analytics has 2/3 market share, and Adobe analytics has one third market share.

Now, note, we are not observing any revenue or costs cost that's changing hands between vendors and publishers. We're only observing these vendor links. Our measure of concentration, then, which is that Herfindahl-Hirschman Index, or HHI, is just the sum of the squared market shares. And this index is going to be increasing and the level of concentration, so 0 is a perfectly competitive market, and 10,000 points is a monopoly. Because is the relative definition of HHI, if all vendors fall by the same percentage, then the relative HHI is going to be invariant. Next slide, please.

Now, we plot relative HHI over time, and we see the evolution of concentration is the mirror image of the average number of vendors. In particular, concentration rises 17% post-GDPR in the short run, and we think the short run is informative, certainly directionally so. When evaluating a policy that has not been enforced, we think that the period where beliefs about enforcement are higher is more relevant, and we'll provide some more evidence about belief later.

We conclude that the GDPR increases concentration. And the intuition for this is that vendors with large shares have large shares because they provide greater value, whether it be because they have lower costs, deliver greater revenue, or have superior privacy compliance. Whatever

the reason, websites prefer to retain the large vendors when the GDPR purchase websites to reduce data sharing, which is why we see this concentration increase. Next slide.

So in the last slide, I showed you aggregate HHI. But we want to define markets more narrowly, and we do so by using an external categorization based on the type of service that vendors provide. And now we can see that the concentration increases in the top four categories that represent over 94% of categorized vendors. In fact, the largest category is advertising, and here, we see concentration rise 25.3%. And in the next three categories of hosting, audience measurement, and social media, we see concentration is still increasing between 2 and 6%. Next slide.

Now, I want to quickly examine one of our three extensions that illuminate the mechanism for the concentration result. We consider the role of the big two companies-- Google and Facebook-- and there are many associated vendors. As before, with all vendors, we see that HHI rises 17.3%. However, when we exclude the vendors associated with the big two, concentration actually falls 6.2%. So maybe we need to update the old adage that nobody gets fired for hiring IBM to also include Google and Facebook. Next slide, please.

Now I want to quickly illuminate some differences by website that tell us something about the economics of how websites are making decisions under the GDPR. Next slide. To begin, I want to break apart the short-run drop in vendors by characteristics of the sites. Next slide.

For instance, here, I break apart sites by the share of traffic they get from EU users. We can see that sites with between 90% and 100% of EU users, on the right-hand side of the figure, drop a little over two vendors on average in the short run, where aside from sites with between 0 and 10%, the lowest estimate on the left-hand side drop a little over five vendors on average in the short run. We think that this reflects the incentives in the GDPR that place a 4% penalty on global revenue. This means that sites with few EU users have relatively little to gain, in terms of revenue from the EU, relatively more to lose from a penalty on their global revenue. The GDPR incentive then has the perverse effect that sites with the greatest share of EU users do the least to cut vendors.

Finally, notice a discontinuity for sites with 0% EU on the far left-hand side, illuminated in orange. Many of these sites are actually not subject to the GDPR, and therefore, do not need to make changes. Next slide.

Finally, we examine the post-GDPR evolution of vendors in 2018. Next slide.

One of the things we noticed after the GDPR is that the average number of vendors grew slowly in countries like Denmark and the Netherlands, but grew rapidly in countries like Bulgaria and Poland. Now, the GDPR is meant to harmonize regulation within the EU, but it's still enforced, in part, at the country level.

So we found a survey measure from the EU that measures regulatory strictness specific to data protection. We found that regulatory strictness is negatively correlated with the post-GDPR

growth in vendor use. This suggests that site beliefs about the probability of GDPR enforcement help to explain the 2018 evolution in vendor use. Next slide-- my last slide.

We start out with a theorized tension between privacy and competition policy, and today, we're able to show you the first empirical evidence of this tension. The GDPR had its intended consequence of decreasing web-technology vendor use and its associated data sharing. But it had two unintended consequences. First, we saw an increase in vendor concentration, and second, we saw that sites with the most EU visitors reduced vendors the least, an apparent side-effect of the GDPR's [INAUDIBLE] design. Thank you.

DANIEL WOOD: Well, thank you, Garrett. Our third presenter is going to be a Jeff Prince. Jeff-- Jeff, you have to unmute.

PRICE: Thank you. There. I'm unmuted now. Perfect. Even better.

So thank you again to the FTC organizers for the opportunity to speak and for for moderating this session. This is joint work with Scott Wallsten at the Technology Policy Institute. And we receive financial support from the Inter-American Development Bank for this work.

So we're looking at how much is privacy worth around the world and across platforms. Next slide, please. So prior speakers have already kind of highlighted this with the GDPR. But this is across many countries around the globe. Governments around the world are grappling with data privacy policy.

And as economists, we're always thinking about the trade-offs of policy. So at a very rough level, we can think about balancing privacy preferences for the citizens with the benefits from use of the data. And one thing that has been emphasized in many places is that it's particularly difficult to measure the privacy preferences. And that's something we're trying to get at with this project.

Next slide, please. So what we do is we use conjoint survey techniques to measure the willingness to accept for online data information. We compare and contrast those willingnesses to accept for a range of data types across different countries, six of them-- Argentina, Brazil, Colombia, Germany, Mexico, and the United States. And then we do it within four different platform contexts-- so with your bank, with your carrier, with Facebook, and then with your smartphone.

Next slide, please. And so for the surveys that we put out, they offered choices with different levels of data privacy and with different monthly payments associated with those different levels. And we have specific reasons for doing it this way. This is rooted in real data markets.

So one example, if you go to datacoup.com, you can go and get money right now for your data, so easy money hanging out there for all of us. So you can offer up some of your data, and they'll give you monthly payments for access to that. So the goal of designing our surveys was to try and make them as realistic of an actual choice that someone would make with regard to their privacy. And the design of these surveys as well suited for measuring trade-offs, which is what we're interested in. Next slide, please.

And so here we have an example for Facebook, just to give you a sense of what we're talking about. So here a respondent is presented with four different options. And each option has different information that's being shared and then monthly payments associated with it. And so to give you a very clear example of a trade-off, if you look at option two versus option four, the information that's being shared is the same except for option two, you're not going to be sharing your texts. Option four, you will, but with option four, you get paid more.

And so it allows people to make the trade-off with, you know, is that additional money worth it to me to give up that information or not? And so then that choice helps us to pin down how people value different types of privacy. Next slide, please.

And so we used the firm called Dynata. When we used it, they were referred to as Research Now. They administered these surveys online for us. We had 325 surveys for each type, and a type being a country, platform.

And then we also included a randomized prompt, where you either got it or you didn't. And the prompt essentially indicated to people the value of sharing their data. And so we included this to try to get a sense as to how flexible people's preferences were. So are they malleable to being prompted about the value of data or not? And so I'll speak to those results as well when we get to them.

Then we had basic screenings on age, so they had to be adults. And then they needed to have existing accounts for all but the banks. Next slide, please.

And so for the analysis, we analyzed the choices that people made, choice data that came back through the online surveys using standard conditional multinomial logit, going back to McFadden's utility-based formulation. And the real basic idea is take the willingness to accept-- to get the willingness to accept for data privacy, just take the utility for data privacy, and divide that by the utility of money. And this is going to be measured in monthly payments. So next slide, please.

So the next few slides present some of our results and tables. This one is a very aggregated result. So what we have here is, averaged across platforms and countries, what is the willingness to accept in dollars per month, using a purchase-price parity index, to compare across countries for the different types of data privacy.

And a summary of what we find here is that A, there is a lot of variation, as you can see. The financial information is particularly well guarded, so those are on the higher end. Also, fingerprint information had high WTA, along with text information and contacts. Next slide, please.

This one here has a lot more information. I know it's a lot to process all those bars, but let me just highlight a couple main points from those. One is, if you notice, the orange bar is Germany. That one is almost always the highest one for each of them. I think for 8 out of 10 it's the highest.

Another, I think, broad point to take away from this is that for the rest of the countries, there's a lot of similarity, and there's not a fixed ordering. So it's not as clear that one country generally has higher willingness to accept than others, with the exception of Germany. Next slide, please.

And then here's virtually everything we have in terms of averages broken down across all the dimensions, so across platform and country for the different types of online information. Again, I know it's a lot to try and swim through, but let me just highlight a couple high-level observations. One is, if you look at the bars across the different countries for the different data types, again you see a lot of similarities in even the absolute values but even more so the relative values.

So if you look at the wireless, upper left quadrant there, the red bars are always the highest, the orange bars are almost always the second highest, followed by the brown. And so there's a lot of consistency in the relative preferences for different types of privacy across the different countries. Next slide, please.

So some key takeaways-- overall, what we find is relative values are quite similar across our six countries. And then another set of results that were harder to present in tables but are also in the paper is that, at a rough level the within-country variation-- so if you think about the distribution of preferences for a particular type of online information with regard to privacy within a country-- is quite similar across countries. So how spread the preferences are, the WTA measures are, across citizens within a given country is similar across countries.

And so some key takeaways from those results is that public and private policies may want to offer similar relative protections, at least to the extent that these countries are representative. If you think about tiered protections, where you think about private firms could offer different levels of protection for different prices, those would likely have comparable appeal across countries. And then the distribution of support for public policy is likely to be similar across countries. Next slide, please.

And then as I highlighted, Germany is different, at least within the set of countries we looked at. They're different overall and with financial information in particular. So a key driver of their outlier status is with regard to financial information. There is a very high willingness to accept in terms of giving up information that has to do with financial specifics.

And then also, with regard to the distribution of preferences within Germany, they appear to be the most homogeneous. So the spread of WTAs for different types of privacy is notably smaller for Germany than for the other countries we looked at. Next slide, please.

And then a few other results that I think are worth highlighting that we found. When we break it down across sex, women versus men, the willingness to accept for women was notably higher than that of men for different types of online privacy, often by about an order of two times. If you go across age, for the older cohort versus the younger, the willingness to accept was substantially higher, often by two, three, or even four times as much. Income, though, does not predict the willingness to accept very well.

And then last but not least, with regards to that leading statement I mentioned earlier, the preferences did not seem to be impacted by that, which suggests that they're not easily swayed by prompts that one might put out with regard to the value of data and giving up privacy, or its potential value. People's preferences seem to be unimpacted by that.

And with that, I will conclude. Thank you very much.

DANIEL WOOD: OK. Thank you, Jeff. Our last speaker is going to be Christine Utz. Christine.

CHRISTINE UTZ: Thanks, Dan and everyone at the FTC, for having me back here at PrivacyCon. What I'm going to present today is a direct follow-up to our GDPR paper from last year's PrivacyCon. This is joint work with my colleagues Martin, Sascha, Florian, and Thorsten and was previously published at ACM CCS 2019.

So I'm sure you've all-- next slide, please. I'm sure if you've seen all of these before. These are consent notices, colloquially known as cookie banners. And these are little pop-up boxes that show up on lots of websites these days. And they inform you of the website's data collection practices and ask you for your consent.

The legal foundation of these notices is the privacy directive of 2011 from the European Union. But especially after the GDPR enforcement, they've seen a large surge in prevalence across websites. And as I presented in last year PrivacyCon, we saw an increase of about 16% between January 2018 and after the GDPR enforcement date.

Recently, there have been some vendors of third-party constant libraries that have started to also implement the new CCPA Do Not Sell requirement in these consent notices. So maybe we'll be seeing more and more of these also with the CCPA as a legal foundation.

Consent notices can be arbitrarily complex. So you can have just the basic one with just an Allow button, like the dark one. Or you can have a more fine-grained selection, where you can select different categories of cookies, like in the one on the top right corner. Next slide, please.

So we saw all of these notices become more and more complex. And we can come up with a couple of research questions. So how often do people interact with these notices? Do different changes in the parameters of the user interface of these notices influence what decisions users make? And why do they choose to interact and not interact with these notices? And what do people expect to happen when they allow or deny cookies? Next slide.

So we decided just to find answers to all of these questions in the field. So we had the opportunity to team up with a German e-commerce website. And that site has about 20K unique visitors per month. Most of them just google something, and then find an article on the website, read it, and then leave the site.

It runs on WordPress and uses common third-party services, like Google Analytics, or embedded YouTube videos, or a design framework called Ionic. And we modified a WordPress plugin to

display arbitrary consent notices on that website. And with this plugin, we conducted three iterative experiments between November 2018 and March 2019 in a between-subject study.

Before we could get started, we had to evaluate the available design space for the UI of consent notices. So we luckily still had a couple of consent notices laying around from our paper from last year. So we just sampled 1,000 of those and inspected them and identified the design space for consent notices. Next slide, please.

So we identified eight different UI parameters of consent notices. Three of them are about the relation between the notice and the website, like the position of the notice on the website, or the size, and whether or not it blocks access to the underlying website. And the other parameters are about the notice itself. So you have the text, whether or not it contains a link to a privacy policy, the general formatting, and then you have the choices offered by the website.

And this is also where notching and dark patterns come into play, because often the available choices are not presented sort of equivalently, but some of them are highlighted, usually what the website owner once users to click, like Accept or Allow Cookies. Next slide, please.

Once we had identified this design space, we designed our study. And in our first experiment, we evaluated the influence of the position of the notice on the website. In the second experiment, we looked at the influences of the different choices offered by the website and nudging. And in our third experiment, we tried to identify the influence of the presence of a privacy policy link, and whether or not the notice uses technical language.

By technical language, we mean things like "This website uses cookies" versus "This website collects your data." Next slide, please.

Our study setup looked as follows. So the user visited the website, and then they were shown one of n constant notices, with n being the number of notices in the current experiment. Our plugin then would log all interactions between the user and the notice, such as clicking OK button, or ticking a checkbox, or clicking the privacy policy link.

If the user chose to interact with the notice, we replaced the content of the notice with another notice that mentions that this is a university study and gave them the choice to either participate or close the notice for once and for all. And if they chose to participate, we just redirected them to our survey.

But we were also interested in people who chose not to interact with the notice. And for that, after 30 seconds without any interaction, we automatically replaced the notice with the study invitation. And then again, if they chose to participate, they were sent to another version of the survey. Next slide, please.

So these are the results of our first experiment, location. In this experiment, we displayed in the binary notice to encourage user interaction. And we displayed it at six different positions. Here and in the following, I'll only report interaction rates. In the paper, we have a more fine-grained analysis that shows what people actually click.

And here we found that the position that yielded the highest interaction rate was in the lower left corner of the screen. We had some theories where this might be the case. So the theory for top versus bottom was that on top, usually the banner is more likely to cover some less important parts of the website, like some header or a menu, while on the bottom you usually have some content and text.

And the same argument applies for left versus right, because if you have texts written in the Latin alphabet, everything is skewed to the left, so more important information can be-- so there's more information on the left versus on the right. Next slide, please.

In our second experiment, we looked at the different options offered by the website and the influence of nudging. And here we had five different banners in terms of option. So we have one banner that doesn't offer you any type of option at all. The one on the very left, you just have a little x to make the notes go away. Then the next banner is one that just has an Accept button. And here you can see the non-nudging variant of the banner. So this Exit button is not highlighted.

The next banner in the middle is the binary banner you've already seen. And here this is the non-nudging, where each button looks like the other. And then we had some more fine-grained options, one that allows you to check or uncheck different categories, and one that has the same for different third-party vendors. And here the non-nudging variants don't have pre-ticked checkboxes, while the nudging variant would have pre-ticked checkboxes.

And here we saw a big influence of nudging. Because in almost all cases, the nudging variants yielded higher interaction rates. And the binary banner had the highest interaction rates. But combined with the qualitative data from our survey, we also saw that the category-based banner was also popular with users. Next slide, please.

We then took a brief look at what people actually clicked. And here's just a quick example. So these are the selections people made on the window-based banner. And you can see if you do not pre-tick the checkboxes, then there were only about 1% of visitors that actively opted in for one of the vendors, while in the case where you have pre-ticked checkboxes, we had about 10% who agreed to data collection by these third parties.

As for the results of experiment three, I don't have the most slides here, because we didn't see any significant influence of either the presence of a privacy policy link and technical versus non-technical language. Next slide.

So then we took a brief look at what people wrote as answers into our survey. And we asked them why they chose to click the banner. And among the reasons for not clicking-- for clicking, excuse me, we saw that one prominent reason was the expectation that the website would not work otherwise.

And this misconception was also present in other questions, like-- next slide-- when we asked what users expected to happen if they clicked Decline or Accept. The top reason or the top statement, what would happen if they hit Decline, was that the website cannot be accessed. And

this was named more often than just mere functionality limitations, which would be much more likely than the website not working at all.

So what did we learn from all of this? We saw that the interaction rate is mainly influenced by position of the banner on the website and not the effects of nudging and preselections. We saw that users appeared to favor a binary category-based approach versus more fine-grained ones, like vendor-based approaches.

And there are widespread misconceptions about how consent notices work. So one of them was that the site cannot be accessed without consent. So one recommendation here would be to inform users about the functionality limitations they can expect when they do not allow the use of cookies.

And then from the survey, we also saw that people have some privacy by default expectations. So they expect no data being collected before they actively make a decision. And this is really not the case in reality. So this would be an issue that could be addressed by regulators, because currently there are no incentives for companies to actively protect the visitors' privacy. Thanks.

DANIEL WOOD: Thank you. I'd like to thank all the panelists for excellent papers and really interesting research that they've contributed to this panel. Now [INAUDIBLE] the way you would through email. And hopefully the slide that's available now is showing you where you'd send them.

But before that, I have a couple of questions for the panelists myself. And I'll start with Christine. So Christine, your research found that the possession of dialogues, that set of choices offered, and other nudges significantly influenced consumer consent choices. How pessimistic should these findings make us about the possibility of mandating the firms that came-- about the possibility that-- mandating that firms obtain informed consent in other online contexts? Or those mandates going to be had to make workable?

CHRISTINE UTZ: I would say that this depends on what this mandate looks like. Because in the case of the EU and GDPR, there was a lot of confusion about what informed consent actually means, because initially there was a big lack of guidelines how consent should be collected and what's actually free and informed consent. And only recently the EU has put out some documents that give a little more insight in that. But there are still lots of questions to be answered.

And I think, yeah, if you want to introduce new mandates for firms to collect online consent, then there really should be some guidelines, along with a mandate that would help companies and anyone else who's collecting personal data to comply with the new regulations.

DANIEL WOOD: Interesting. I guess I'll rotate through the panelists. And my next question is for Jeff. So the privacy paradox, roughly stated, is that people report stronger preferences for privacy in surveys than they demonstrate with their actual behavior.

How well do the valuations for privacy you recovered in your research, Jeff, match up to valuations from consumer choice-based studies?

JEFF PRINCE: That's a great question. I think the closest to us prior to our work was Savage and Waldman did some measures for the value of privacy with regard to apps. But there are some distinct differences. So I mean, obviously those are single apps. They looked at one time payments. And theirs were kind of in the $1 to $4 range.

So are ours a lot bigger? I mean, technically yes, since ours are monthly payments. But again, you know, we're looking at major platforms rather than a single app. So people might look at that decision differently.

More broadly, I think, you know, with the privacy paradox, it's tough to say. I think, you know, this is one of the reasons why I think more quantification is valuable, because a lot of times, as people know, we ask people, do they value their privacy, and the answer is yes, maybe even a lot. But then that might not line up with what quantifiable metrics would be in terms of how much they value their privacy.

And in fact, even with our numbers, we had outlets interpret our numbers as being large, and we had outlets interpret our numbers as being small. So in some ways, it's a lot about perspective. So it's hard to align our figures with what people and people's qualitative analyses have revealed. But I think, you know, our hope is to contribute more to the quantifiable version of people's value for privacy.

DANIEL WOOD: Right. Well, so it seems to me like your numbers are going to become a standard for [AUDIO OUT] in the privacy space. So I thought they were really interesting.

JEFF PRINCE: Thank you.

DANIEL WOOD: My next question-- like I said, I'm rotating-- is for Garrett, Garrett Johnson. So you talked a little bit about this, but it seemed like a lot of what you measured was the short-run adjustment to GDPR enforcement. Should we evaluate the effect of GDPR based on that? Or is long-run adjustment what we should be interested in, or are both useful?

GARRETT JOHNSON: There we go. It's an important question. My answer is emphatically that the short-run provides the best available evidence. But I want to make the three big points here. The first is that the GDPR is really confusing to study in this industry, because regulators have been slow playing the industry.

So even as of today, the EU has not fined any of these websites or technology vendors. But at least three EU countries have criticized the industry for practices that do not comply with the GDPR. So again and again, the EU keeps delaying enforcement and giving the industry time to adjust its practices.

So the big question is, how are you going to study a law that hasn't been enforced? So we try to argue that the best time to do so is when the firms are most afraid of enforcement and change

their behavior accordingly. And our evidence does suggest that police plays an important role, like the fact that countries that face stricter regulators seem to keep their vendor use lower than those that don't-- and also, some evidence that also we talk about in the paper.

And the last thing I'll say is that this is an industry that moves very quickly, that has-- it's a fast growing market. And we see in general that the number of vendors increases over time. So given this fact, I think the best evidence we have is this big trend break we see right around May 25, 2018, which is the month right after the GDPR deadline.

DANIEL WOOD: OK. Thanks for the clarification. So again, let me encourage questions from the audience. If you want to ask a question, email privacycon@ftc.gov, and we're very interested in your questions.

Let me ask a question of Guy in the meantime. So Guy, how do you think your results will generalize to other domains on the internet, beyond online travel?

GUY ARIDOR: Thanks. Yeah, I think it's a good question. So when thinking about-- I think there's two things. So the first is that I think there's two dimensions to think about from the consumer perspective. The first is thinking about in a particular context what is the instrumental value of privacy and how to consumers perceive firms using their data. So in the context of online travel, at least anecdotally, consumers are more likely to be privacy conscious than they may be in other settings, because they think that consumers are-- or that firms are making use of this data.

And the second is that a lot of our analysis sort of hinges on understanding the impact of consumer histories. And so for instance, if you compare this to a setting, like social media or something, where you see consumers much more often, it might not be as applicable. But in settings that are similar on those two dimensions to ours, we would expect the results to generalize. We would expect the sort of externality results to generalize to any setting.

Finally, I think it's also important-- this is something that we spent some time trying to grapple with trying to contextualize our results in the context of the broader advertising ecosystem. So in particular, like the-- you know, the advertising partner-- the intermediary we partner with sort of views itself as a competitor to Google. And we think that our study sort of helps understand how a niche advertising intermediary gets impacted in terms of profitability in data observed by a smaller advertising intermediary.

And we suspect that advertising intermediaries in other domains would be similarly impacted. And I think Garrett's paper sort of points to this, and there's a few papers that are pointing to the need to sort of think about the broader competition effects of GDPR on these things. And so while we find that our results aren't wholly negative on the firm side, it would be interesting to think about how that would compare to, say, Google, who might not have been as impacted by GDPR as our advertising intermediary.

So we suspect that there are results on the impact to a third-party intermediary would be similar. And it's interesting future work to think about how that would impact itself in a broader competition between advertising intermediaries.

DANIEL WOOD: Cool. So let me ask you another question.

GUY ARIDOR: Go ahead.

DANIEL WOOD: So part of what I found fascinating about your paper was that there were these externalities between different types of consumers. How is the move from-- can you throw in a little bit more and then tell us how the move from software-based data obfuscation to GDPR opt-out is likely to affect the welfare of consumers who don't have strong preferences about privacy?

GUY ARIDOR: Yeah. So that's a good question. So I guess as an economist, I have a slight caveat that we do a reduced form metrics size, so we can't directly say anything about welfare. But we do argue indirectly in the paper that if you think of consumer welfare as largely depending upon the quality of services they receive, this is largely dependent on the ability of a firm to do prediction.

And so particularly in our context, we find that there there's a marginal improvement in prediction. And this may lead to other domains and better personalization and ultimately improve consumer welfare. There's obviously setting such as where firms are using this prediction to do price discrimination where, you know, arguably it would reduce consumer welfare.

So the way we try to frame it is, the effect for consumers really depend on the alignment of preferences between firms and consumers in terms of how they use their data. But I think it would be-- and again, we point this out the paper-- it'd be interesting to do a proper structural analysis to really decompose the welfare benefits to these policies.

DANIEL WOOD: Cool. Let me turn back to Garrett. So Garrett, while the California Consumer Privacy Act is sometimes compared to GDPR, there are some differences. Do you think the CCPA-- or if you want you could imagine a different hypothetical federal privacy legislation. Do you think that sort of legislation would lead to similar increases in concentration that you find?

JEFF PRINCE: I think it would have different effects on competition. So I think the main difference is that the GDPR has a data minimization principle. And that places pressure on firms to limit their data sharing partners. To my knowledge, the CCPA lacks this principle, which seems to be doing most of the work in our setting.

Instead, the CCPA operates on a notice-and-choice basis. So it basically tells sites you need to put some Opt Out button on their site that allows consumers to avoid data sharing. And we know from research, like Christine's and Guy's and my own works-- that opt-out rates, at least according to our stuff, is like 5% to 15% when sites have to display this prominently.

But the CCPA insists on this colorful language which is, "Do not sell my personal information" for the Opt Out button, which I would speculate-- it would be interesting to hear Christine would

say about this-- I would think this could increase the opt-out rates. So while I don't think this is going to have effect on vendors, I do worry this could have an effect on the publisher side. In particular, large websites may have an easier time gaining consent than smaller and less recognizable websites.

And our work examining over a thousand firms using Adobe Analytics Data and the GDPR is consistent with this finding. We see a larger reduction in recorded web outcomes for smaller websites in our data.

DANIEL WOOD: OK. Christine, Did you want to speculate on the effects of the CCPA's-- what's the exact wording, Garrett?

JEFF PRINCE: Do not sell my personal information.

DANIEL WOOD: Is that a good notice?

CHRISTINE UTZ: Actually, we already did some investigation of that. So we took a look at a couple of-- I don't know how many-- couple of thousand of US websites, and we looked at how they implemented CCPA link. And we saw really, really big variance in how this link is named. Often, it's "Do not sell" and "Do not sell my info," "Do not send my personal info." There were dozens of variants on that.

Yeah. This makes you wonder how the sites will deal with the rest of the CCPA requirements if they already have kind of difficulties complying with this simple requirement like, just put in a link that has this wording?

DANIEL WOOD: Yeah. My understanding-- and I haven't been following it super closely-- was that I think the California attorney general might be producing guidance. I don't know if they have [INAUDIBLE] deadline. But the guidance is-- you know, if they read your paper, the guidance might be great.

So let me ask a sort of very broad question of you, Jeff. What's the most important points about privacy policy that you think privacy politics policy makers should be taking from your research?

JEFF PRINCE: Oh, wow, that is a big one. I guess, you know, one of the takeaways for us was there was a very noticeable similarity in both the relative and, even in a lot of ways, the absolute preferences for the different types of privacy. I think many would have predicted ahead of time that Germany would come out with the biggest numbers, and they did.

Although even with Germany, if you take away the financials they're not that that much higher than everybody else. And we have a pretty wide cross-section. So I mean, obviously we're influenced by our funding source in terms of where we directed our focus. So we had a lot of Latin American countries. But I also think we didn't know a lot about what was different privacy preferences across those countries.

So, you know, that was one of the big takeaways, at least for me, is that when you think about privacy policy-- how people value privacy in a relative sense-- across countries and across different types, there wasn't that stark of a difference across countries, even though obviously there's vast cultural differences and other differences across those countries.

DANIEL WOOD: Yeah, that was interesting. So it seems like-- just throwing this out-- it seems like that sort of structural similarity makes me a little more confident in the survey-based approach. It seems like it might be producing something, the real preferences that we're finding consistency across different countries and across different groups of people.

But let me turn to back to Christine. So Christine, you found that some users have misconceptions about how either they were-- how either what interacting or not interacting with consent notices would tell the website how they should use their cookies. How can we improve-- what's the best way to help these users have more accurate expectations about the very simple act of interaction?

CHRISTINE UTZ: Yeah, that's really an interesting question. And we've been wondering that, too. So I think one big step in the right direction would be to actually make websites comply with the user selection. Because there are other papers that have shown that many web sites already start tracking before you've actually made a decision.

And the next step would be to just tell them which parts of the website they can expect to work and which won't work if they decline certain types of data processing. Like, for example, in the case if you have an embedded YouTube video, you can say, OK, if you don't agree to YouTube setting cookies, then you just will see a gray box or something and not the embedded video, something like that.

DANIEL WOOD: OK. So we did get one-- actually, two audience questions. And one of them is about Privacy Shield. And that's a big area, so we're not going to touch that yet.

But the other one is a somewhat more specific question for Christine. So we only have about a minute left. But Christine, if you feel like you can do justice to this, how can data protection authorities who often review consent statements for consent but not other factors incorporate your research into their day to day work?

CHRISTINE UTZ: OK. Yeah, I think they could maybe feel encouraged to issue, to come up with some guidelines. I mean, we already have some guidelines. One was recently published by the EU. And we have some other guidelines by different EU member states. But I'm sure there are still lots of uncertainties what constitutes valid consent.

And then one big issue we still see-- this is something Garrett has already pointed out, which is we do have the laws, but there's just a big-- they're not being enforced right now, or just a very, very small extent. So right now, there's really no incentive in many areas to comply with GDPR, because there's just a lack of enforcement.

And one big problem in this area is that data protection authorities just lack the funding and the personnel to actually enforce the law. But I hope we'll see some changes in that in the future so that the regulations can finally exist, not just in paper, but also actually in live systems.

DANIEL WOOD: OK. Sounds pretty simple to me-- speaking purely for myself and not for the Federal Trade Commission. I would like to thank you all again for a wonderful panel and for participating in this year's PrivacyCon. PrivacyCon will resume in about eight minutes. But for now there's a virtual coffee break.

And when we resume, we'll do the last panel of the day on miscellaneous topics in privacy and security. So thank you. Thank you again.

GARRETT JOHNSON: Thanks, Dan.

JEFF PRINCE: Thank you.

CHRISTINE UTZ: Thanks.

JAMIE HINE: 3, 2, 1. Hi, everybody. Welcome to our final panel of the day, panel 6. This is sort of the miscellaneous panel. We'll call it the potpourri panel for today on privacy and security issues. Just a couple of reminders-- one that you can send questions to the privacycon@ftc.gov mail address.

And if you're tweeting, please make sure to use the hashtag #PrivacyCon20. So we have four great panelists today to round out the day. We have Hana Habib from Carnegie Mellon University; have Ido Sivan-Sevilla from Cornell Tech; have Daphne Yao from Virginia Tech; and we have seen Yixin Zhou from University of Michigan's School of Information.

So without further ado, we'll turn the floor over to Hana Habib from Carnegie Mellon.

HANA HABIB: Good afternoon, everyone. Today I'll be presenting two research papers on the usability of online privacy choices on behalf of my co-authors at Carnegie Mellon and the University of Michigan. These papers are published at the Symposium of Usable Privacy and Security 2019 and CHI 2020. As I mentioned, our focus in this work are privacy choices on the internet.

And the next slide has examples of regulation which mandate these privacy choices. And this includes the GDPR in the European Union, as well as the CAN-SPAM Act, COPPA, and now the California Consumer Privacy Act in the US. Additionally, groups like the Digital Advertising Alliance also work towards self-regulation in the advertising industry.

On the next slide are examples of three types of privacy choices that are commonly mandated by regulation and self-regulatory guidelines. And these include opt-outs for email communications, opt-outs or targeted advertising, as well as data deletion choices.

Next I'll go over our research questions, which explore how these mandated privacy choices are provided in practice. We asked, what choices related to email communications, targeted advertising, and data deletion websites offer? Additionally, how are websites presenting these privacy choices to their visitors? And what are the potential usability issues?

To answer these questions, we conducted two studies. The first was a manual in-depth content analysis of privacy choices on 150 websites. We followed up on this work by conducting an in-lab usability study of a subset of these choices.

So next I'll go a bit into more detail about our study protocols. To standardize the data recording for empirical analysis, for each website we filled out an analysis template with 82 questions. An example of these questions included the location of the privacy choice-- was it in the privacy policy, account settings, somewhere else on the website; the level of detail provided about each choice; the availability of links to the choice; as well as the path of implementation, for example, how many user actions were required to actually use the choice.

So next I'll provide a quick overview of how we selected websites for the study. We randomly sampled 150 websites from Alexa's Global Top 10,000 list as of March 2018. All 150 of these websites were analyzed between April and October 2018, and half of them were reviewed by two researchers with an agreement of 0.82.

Next, a bit more about our user study. For our user study, each study session we conducted had three portions. First, we conducted a pretest interview to understand what users already believed about data collection and privacy controls.

Next, we had participants complete to study tasks. First study task, we identified a set of nine websites that had common implementations of privacy choice mechanisms that we identified in our empirical analysis. We gave users scenarios to describe this privacy choice task and asked them to complete this task as they would in the real world.

For some websites, the scenario would require going to the account settings, while another user required going to the privacy policy. And the policy mechanisms could appear as links in the policy text, or it could be described within the text as instructions. In the final component of the study, we asked participants interview questions after the test to capture information about their experience and understanding of the study tasks.

So next, I'll go over a few of our results. But I encourage you to read more in our papers. So first, some good news. From our empirical analysis, we found that privacy choices are common. Almost 90% of websites that use email marketing or targeted advertising in our sample offered their respective opt-outs. And almost 3/4 of all sites in our sample provided a data deletion mechanism. Next slide, please.

In our empirical analysis, we found that privacy choices were often provided in privacy policies. The downside of that, other than consumers largely ignoring privacy policies, is that the headings under which choices are presented are inconsistent from policy to policy.

This table presents bigrams and trigrams in headings of sections that describe these privacy choices. We noticed that some terms were evenly distributed, like "your choice." However, there were more unique terms for certain types of choices, such as opt out for email communications, third party for targeted ads, and your right for data deletion. Alarmingly, no single n-gram occurred in more than 20 of our analyzed policies. And this lack of consistency across websites could make it hard to locate choices in privacy policies.

Next, I'll go over another reason why offering choices through privacy policies is less than ideal. From our user study, here's an example of a privacy policy that users encountered on one of the websites in the study during the scenario in which they were trying to stop seeing ads for shoes that they searched for last month. So here are some relevant information about how ad partners use cookies and beacons to decide which ads to show.

On the next slide, we see that the first link here is for opting out of Google Analytics. And participants often clicked that first when trying to disable cookies. But this link isn't that useful if the main goal is to disable cookies, so it's not clear why it's shown first here. On the next slide, we see that the information about disabling cookies was presented underneath that link. Next slide, please.

So from our empirical analysis, we noted that another reason why figuring out what to do could be difficult is that websites sometimes provide multiple tools for the same type of privacy choice on different pages of the website. So take Twitter's targeted ads for example. First, in the account settings you can find the opt-out provided by Twitter itself. If you navigate to it's About Ads page, it only shows opt-outs provided by the DAA, NAI, as well as Google. If you go to its privacy policy, only the ones provided by Twitter and the DAA will show up.

All of these links to multiple opt-out tools spanned across multiple pages of the website may cause confusion about what tools should be prioritized and what their differences are. In fact, this is something that we observed in the lab. On the next slide, we have an example of a privacy policy participants saw in one of their tasks.

Participants who saw this had a difficult time understanding which of these three links would allow them to opt out of targeted advertising. While it was confusing when there were multiple links leading to different tools, when there were multiple paths to the same choice or information related to a privacy choice, we observed that it actually tended to be easier to find. On the next slide, we have an example from the lab.

Most participants who were assigned a data deletion task on runescape.com found the information they needed through searching the website support pages rather than referring to the privacy policy. And this led them to the Your Personal Data Rights page shown on the next slide, where they were able to see that the website offered this.

Another major result, which I present on the next slide, is that using these choices require high numbers of user actions. The user actions could include clicks, hovers, scrolls, filling out form fields, or other types of interaction. For example, we see here that on average a participant took about 38 actions to exercise a privacy choice using a policy link.

And this average includes the reality that most users make some mistakes, like going to the wrong page, clicking the wrong item, on the way to the final correct action. When we collected data about the shortest path to each choice in our empirical analysis by performing the same tasks with prior knowledge of the location of the choices, it still required a high number of user actions. In the case of policy links, even if someone already knew exactly how to get to the final step and took the shortest possible route to get there, it would still take about 22 actions.

In the lab, we uncovered some practices that required unnecessary effort. On the next slide, here is an example of a part of the complicated form that some websites require to exercise a privacy choice. The example shown here is a form for deleting data on the New York Times website. Most participants dislike the number of similar-seeming options here. Further down the form, you had to select from a list of 22 different New York Times services, and you could only submit one request type at a time.

The next slide provides another reason why exercising privacy choices might require unnecessary effort. So websites sometimes require users to submit written requests to complete actions, such as data deletion, when a simple web form would have sufficed.

There were also participants who ended up writing emails to customer service to ask for help, because they couldn't find a simpler way to do their task through the website itself. And it sometimes wasn't easy for participants to articulate what they wanted to do. For example, one participant who was given the shoe ad scenario I described before wrote this email to ask for help.

So after hearing so many issues, you might wonder, how do we improve the usability of website privacy choices? On the next slide, we show that one way regulation could help improve visibility is to have explicit requirements that dictate parameters like the location of controls and the way that controls are presented. And the findings from our user studies suggest that the CAN-SPAM Act has likely been effective in making email unsubscribing more usable. It mandates the look and placement of email opt-out links in commercial emails, and users thus expect to find the Unsubscribe link in that location.

Additionally, the next slide shows another way that policy could play a role, which is by standardizing policy section headings so that choices are easier to find. Such practice has been adopted by the US financial industry as a model privacy form to help financial institutions comply with the GLBA. Though it may not be perfect, it's definitely a good start. And research has shown that the standardization effort of the GLBA contributed to less ambiguity in privacy policies.

As summarized on the next slide, another way that choices could be made more usable is through unified settings. This could simply mean matching user expectations by always having privacy choices easily accessible within websites' account settings, rather than buried elsewhere on the website or its privacy policy. There is also the possibility of further unifying choices for users by offering more universal mechanisms, such as through a web browser that's able to parse privacy policies or use machine-readable privacy policies to help users exercise preferences across multiple websites with less effort.

Some of our group's recent research in the context of the California Consumer Privacy Act has also explored unified visual standards to help users find privacy options on websites. The next slide provides an example this. The Privacy Options button here on the bottom right that we've tested is designed to convey the idea of choice and to serve as a central location for all privacy-related choices on the website. Ideally, this would lead to a dashboard that could also interface with automated tools to allow users to control privacy settings across multiple sites.

On the next site, I wanted to quickly recap our work. We conducted an empirical analysis and in-lab usability evaluation of email opt-out controls, targeted advertising controls, and data deletion mechanisms. We found that privacy choices are prevalent but suffer from several usability issues. And our findings suggest that the standardization of choices through regulation could improve usability.

For more information about our ongoing work, feel free to visit this URL. I'd also like to give a shout-out to my colleagues in the Usable Privacy Policy for their contributions to the research, as well as our funders. Thank you, and I'd like to pass it off to the next speaker.

JAMIE HINE: Thanks. And that will be Ido Sivan-Sevilla.

IDO SIVAN-SEVILLA: Right. Thank you, Jamie. Good afternoon, everyone. I'll be presenting our research today about the extent that third-party trackers in websites persistently identify users across websites, or more specifically across social contexts. And this research was conducted with the research group in Cornell Tech, including Wenyi Chu and Xiaoyu Liang, two Cornell Tech master's students, and Professor Helen Nissenbaum, Professor of Information Science in Cornell Tech. And we gratefully acknowledge support from NSA and NSF for this research.

Next slide. OK, so a little bit of background. If you think about the web, the web is an array of different social contexts. We go to the web when we want to look for information about our medical problems or express our educational aspirations or consume news. And advertisers take advantage of the fact that the web is an array of all these different things to conduct cross-context inference about individuals.

The fact that the web is an array of social contexts coming together is really profitable for this industry. Think about, for instance, how advertisers can cross information about users' medical problems, educational interest, and news consumption habits. They become in a better position to know when a consumer can be turned into a purchaser and make purchasing decisions.

One more click. And the fact of the matter is that we are never alone in the web. Embedding third-parties in websites became an inevitable and disturbing social norm. According to recent statistics, there are nine trackers on average per website and overall 33 tracking requests per page. And these trackers have the potential to undermine the integrity of our context and the way we browse the web and violate our privacy according to our approach of privacy as contextual integrity. Next slide, please.

So this approach was also used in a previous paper presented by [INAUDIBLE] in this conference. And we argue that privacy is the appropriate flow of information based on

information norms in a given context. Privacy is not about control, whether information is public or private. It's about how we use information.

So think about some examples of privacy violations according to this theory. So think about employment decision based on religious affiliations. Think about the display of advertisements based on sensitive health information. Think about clinical tagging based on voice assistance data. These are all examples in which information was taken out of its original context without-- against the privacy expectations of the data subject that their information is about. Next slide, please.

So what we're trying to do here is to apply this context-sensitive approach to online tracking. And when you look at previous studies, you see that online tracking was studied in bulk, across thousands of website, without distinguishing their social contexts. And our approach here was to apply a context-sensitive analysis to online tracking-- comparing tracking across different social contexts of the web.

So what does it mean? Let's try to visualize what we're doing here. So one more click.

So for instance, if you go to webmd.com, one of the third-parties you'll see there is doubleclick.net. DoubleClick uses a user ID cookie and assigns you an ID, in this example the string 129. One more click, please. Then you go to nytimes.com, and you see the same third-party tracker. And, one more click, the tracker assigns you-- or it uses the same user ID for you when you go to nytimes.com and can potentially link information about you from both of these browsing sessions. This is the exact cross-context infringement we're talking about.

So in this research, we're trying to label third-party trackers as what we call "persistent identifiers." So among all the third-party trackers out there in popular websites, who are those who persistently identify users across different social contexts against our privacy expectations? Next slide, please.

OK. So a little bit about our methodology. So we used an instrumented Firefox browser based on the Open Development Project from Princeton University to investigate the top popular 15 web sites in three different contexts-- in health, education, and news contexts. One more click.

And as you can see from our chosen websites, these all embed very different dynamics or interactions for the users. Popular news websites, we consume news, we express our interest in news articles. For the health care context, we might express or share some information about our medical problem that we expect this information to be kept private. And finally, in educational context we express our educational aspirations and maybe hint on our future career goals.

And we conducted six different experiments according to the different possible browsing sequences between these three different social contexts to realize whom among these third-party trackers persistently identify users across these different contexts. And it's important to remember that what we argue what matters here is not only the amount of checking within a website but also how those trackers choose to persistently identify users across the social

contexts. We expect our information from the health care context to be used for health advice rather than for commercial purposes in other websites.

According to user surveys, and some of them were discussed in previous sessions in this conference, people are not comfortable with trackers navigating data between different contexts to get a better understanding of their profiles. And this is what we're trying to measure in this study. Next slide, please.

So a better data analysis approach. So for each experiment, we were matching ID cookie among the contexts to realize which trackers use the same user ID for every context. So first, we observed all the third-party trackers that interacted with our browser. Then we detected all associated cookies of each tracker and grouped our data based on cookie name and cookie value pairs. Then we selected identical cookie values that appear in more than one social context.

And finally, we applied a known methodology to recognize among all these cookie with an ID cookie based on the uniqueness of this cookie and its length in the browsing session. We did not simply assume that the presence of tracker in two different social contexts means that they persistently identified the users across those contexts. Instead, we looked for valid evidence, in this case the usage of the same cookie ID across these contexts, to assume this persistent identification trend.

And we acknowledge that our results represent only a lower bound of these persistent identification instances. We are aware that cookie values are often hashed or encrypted when used by the same tracker. We also acknowledge that persistent identification of users happen in server side as well, in ways that are more challenging for detection. So we expect our results to be considered as the lower bound of the amount of persistent identification that's actually happening in the web. Next slide, please.

OK. So now let's see some overview of our findings. Ultimately, we found that social contexts matter for trackers. We found a third of the studied third-party trackers use persistent identifiers among all three social contexts. Secondly, we saw that the three contexts that are linked by third-party trackers are linked to a different degree based on the website that is under study. And I will show this in a moment.

And finally, and maybe most interestingly, we found that third-party trackers are more likely to persistently identify users following users' visit to health care websites. And this is especially alarming in our times of the global pandemic, when health care websites are becoming extremely popular, when users seek health information. Next slide, please.

OK, this figure is trying to start and capture and present what we found. So we've overall found that user IDs that were generated while browsing in health care websites are more likely than others to follow users to other social contexts of the web-- to news or to education contexts. And here you can see that 68 of the third-party trackers, when we visited health care websites first, were labeled as persistent identifier.

It means that the user ID that was generated for you when you visited a health care website is likely to follow you by more trackers than in other experiments when health web sites were visited after different websites. So the user ID that was initiated for you when you visit the health care websites is highly appealing for third-party trackers in other social contexts of the web. Next slide, please.

And this is a complementary figure that shows that for news websites-- so when you visit news websites after health websites, you see 69 persistent identifiers-- 69 third-party trackers that are turning to persistent identifiers linking the ID that was assigned to you from health care website. So this is very appealing for third-party trackers that operate in other websites to know that you are visiting health care websites as well. Next slide, please.

Then we decided to zoom in. And we wanted to graph how this is happening between the websites, how persistent identification works between all the different websites that were under investigation. So we created an edge between two websites in case an ID cookie was used by a tracker that is present in both contexts. In this example, we saw in WebMD and nytimes.com, Twitter, which is in this case the third party, used the same user ID in both of these contexts, potentially linking our browsing habits from both of these contexts. One more click.

And if there was more than one third party, our edge became thicker. So we're trying to understand the scale of this persistent identification trends between websites. Next slide.

OK, so how this looks when you look at scale at all of the websites under investigation? And I know this might be a little a little tiny. So you can zoom in on the upper right corner of your screen to get a better look of what we're trying to visualize here.

And what you see here is our browsing session moves from the health care to the news to the education context. You see that third-party trackers in news websites link user ID from health websites, potentially violating our expected privacy norms from these websites. So you see that for each and every health care website, there is a link. To a varying degree, we have persistent identification trends to a news website. So this is very appealing for trackers in news website to link our ID and study about our behavior from these health care websites. Next slide, please.

And here you can see this again. We're moving from health to education to news website, and you see the dominance of persistence identification from health care websites. Trackers from each health care website identify user in each of the education websites and in the news websites, but in different volumes. So the thickness of each edge means that different number of trackers are actually following this trend of persistent identification. Next slide.

OK. Three takeaways from these studies. So first, we see that users who consume their news or visit educational resources after browsing at health care websites are potentially more vulnerable for manipulation by the advertising industry. Like I said, this is especially alarming in times of the global pandemic.

Secondly, what matters for users' privacy is not only the amount of tracking within a given context, but also the extent that trackers link information about users between those contexts for

potentially better targeting purposes. And finally, like I said, health care website, which were regarded in previous studies as less dangerous for users' privacy because they had less number of third-party trackers that were following you, are actually the most alarming ones when it comes to persistent identification trends.

And next slide, that will be my last one. So to conclude, we argue that this is a first modest step to apply contextual understanding to online tracking. We argue that this is a rather unaccounted privacy violation. We should all work for keeping the integrity of our different social context when we go online, no matter how profitable their conflation may be for certain parties, in this case third parties and advertisers.

And ultimately, this work is a call to apply a more context-sensitive analysis to online tracking in order to better understand this rather unaccounted privacy violation. So more, of course, is in the paper. And I'm looking forward for your questions and comments. Thank you.

JAMIE HINE: Daphne, you have the floor.

DAPHNE YAO: All right. Thank you, everyone. Thank you, Jamie. I'm Daphne Yao from Virginia Tech. Of them than the talk about payment card security. And this is work published in ACM CCS 2019 last year. And it was in collaboration with my PhD student Sazzadur Rahaman, who just defended his PhD thesis yesterday, and my colleague Gang Wang from University of Illinois. Next slide.

PCI stands for Payment Card Industry. The body behind the data security standard, the DSS, are big banks-- Visa, MasterCard. They form what is called as the security council, PCI Security Council. The standard, it started in 2004, many years ago.

A little bit of history. Before there was DSS 1.0, Visa came up with data security standards on its own. And quickly, many other companies-- MasterCard, Discovery, American Express-- followed suit.

And then it was so confusing. The payment ecosystem had so many intertwining components. The acquiring banks, the issuer banks, the merchants, they have to work together on transactions. And so it's very confusing to have one standard for Visa and another standard for MasterCard.

And so the big banks have formed the Security Council and decided that let's just unify all the data security standards. The current version is 3.2.1, which has evolved tremendously since its first version. The 4.0 version will come up in 2021.

So I got very interested in PCI-- next slide. So I got very interested in PCI DSS because of Target data breach. I wrote an article explaining the details of the Target data breach. It occurred in 2013. And some of you may know the initial entry point of the attacker was this air conditioner system. So [INAUDIBLE] mechanics one of the employees there fell victim to a phishing attack.

And eventually, that person's credential was used to access internal Target networks because of the lack of network segmentation. And eventually, malware what is called the BlackPOS was installed on point of sale devices in Target. 40 million credit card numbers were compromised.

So as I was reading about Target data breach, Target was actually in compliance with DSS, the Data Security Standards, back in 2013. And that was one of the main arguments that Target's CEO Gregg Steinhafel used to say, oh, we are in compliance. We got breached. It's not our fault.

But as you look into the standards, you realize that a lot of those measures were just a sanity check. It was just a baseline. So I will explain a little bit more about the exact measurement we did. Next slide.

So as you look into a bit closer about the DSS standard, you realize that regardless of which merchant size you are-- you can be Walmart, you can be mom and pop shop, 7-Eleven-- you have to satisfy this, what is it called, a quarterly scanning report. It is an external scan of your network, the payment network of the merchant, and to ensure that all the system that touches the credit card has to be compliant with the set of standards.

For bigger merchants that have more than 6 million transactions per year, they have to follow additional requirements. For example, the auditor has to be on site and go through some internal design configurations to ensure the correctness and security of the systems. And so we decided to focus on this ASV, this Approved Scanning Vendors, how secure it is. And from a scientific point of view, can we quantitatively measure it? Next slide.

And so this is not really anything that people have done before in a way that has some quantitative measurement of commercial scanners. And part of it is to understand the requirement of DSS specifications and then be able to reflect it in some sort of a testbed that allows you to test the scanners with. Next slide.

And so we spent a long time designing-- you know, how should we set this up? And so we eventually decided to put together what we called BuggyCart testbed. It is e-commerce website. It's a web application that sells electronics. It has a card payment system. It has different options for the user to design their purchases.

And so we used this as a testbed and embed altogether 35 vulnerabilities. But only 29 of them can be scanned externally. So we only need a scanner to find out 29 of them.

And so we find out there's numerous-- more than 100-- scanners to choose from. And of course, from a scientific research group, we have limited budget. But we tried to cover high-end scanners and low-end ones. And luckily, some of them offer free trials. And so we selected a few of them to test.

The way that we tested the scanning services is we just do a baseline scan and see how many vulnerabilities they can find. And then we'll follow their instructions to fix some of them, but then only the minimum amount of fixes. And so eventually, we'll have a version that all of the testbeds indicating the minimum fix, and a testbed that can pass the certification. So next slide.

A quick summary of our findings. I'm going to explain a bit more. Five out of six scanners knowingly certified vulnerable merchant websites. And this is somewhat expected, also somewhat disappointing. And I'll explain more why somewhat expected.

In addition, we also put up our own scanner, a lightweight one. We scanned a whole bunch of websites. A majority of them are not fully PCI compliant. Next slide.

A quick summary of the findings on the scanners. We eventually settled six scanners. Two of them are advertised as two different products, but they use the same engine. Two other scanners, 3 and 6, are not approved. They are not approved ASVs.

If you look at this, the last column is the most important one. Only one scanner, scanner 2, does not allow vulnerabilities knowingly to exist in a certified version, even though there are seven vulnerabilities it cannot detect of the 29. So this is a very disturbing result.

The must-fixes has to be vulnerability score greater than 4.0. And it was defined in the ASV scanning guideline to have to be automatic failure. You have to-- the scanner has to fail them, the website. But most of them don't. Next slide.

More information about the certain type of vulnerabilities called the application security-- and those are the typical cross-site scripting, cross-site request forgery, SQL injection, the harder one, the harder vulnerabilities-- failed miserably. On the right last four columns, those are research products. They are top-of-the-line web scanners. Some are research products, some are commercial products. They also don't do very well. So this is-- give you a serious pass of what's going on here. Next slide.

Good news is that when we use our scanner scanning websites, a majority of them, even though they are not fully PCI compliant, some of the typical issues don't exist-- default mySQL username/password, weak hashing certificates, browseable directory-- those are ../ and you can go back. And those are gone, which is good.

Animation, one click, please. However, we've seen wrong domain names, vulnerable OpenSSH versions, expired certificates. And those are the issues that PCI compliance prevents but that still exist. Next slide.

And of course, that's not very surprising. If you have inadequate scanners certifying insecure websites, you inevitably will have vulnerable websites. And so this is a first quantitative study measuring PCI scanner capability. But then the issue is much, much more beyond the PCI by itself.

We tested web scanners, they don't do well. We tested research product, they don't do well either on certain type of more complicated web application vulnerability. So what does it mean?

And so if you can remember one thing, that's this slide. For our various stakeholders, everyone needs to improve. This is definitely not some work to say scanners-- you should be blamed. No, no, no. Everyone needs to improve.

The research community needs to have more deployable solutions. The cross-site scripting, a concept that's been around for a long time, there's no good open-source deployable grade solutions. Regulatory authorities-- how can we improve the specifications? And then part of it is to have a holistic measurement of system security as opposed to just one check, one check, one check, put them all together. Scanner evaluators-- how to improve, more importantly, more robust testbed. Next slide.

Last slide here. PCI specification, very comprehensive. We were very impressed about the completeness, but enforcement is tough. Research needs to catch up. We also disclose our findings with the Security Standard Council and got positive feedback. And this is a problem that needs everybody in the community to improve.

That's it for my talk. Thank you.

JAMIE HINE: Excellent. Thank you so much, Daphne. Yixin, final presentation.

YIXIN ZHOU: Thank you, Jamie. Hi, everyone. My name is Yixin Zhou, and I'm happy to present this research collaboration between University of Michigan and NortonLifeLock Research Lab on the adoption and abandonment of security, privacy, and identity theft protection practices. This paper was published at CHI 2020 with the Best Paper Honorable Mention, and was sponsored by NortonLifeLock Research Fellowship. Next slide, please.

Consumers need to know how to protect themselves online. Data breaches, hacking, phishing are but some of the threats they face. While there's lots of useful expert advice on how to protect oneself for privacy and security online, study after study shows that most consumers do not adopt best online security practices, potentially leaving them at risk. Next slide, please.

What we don't know, however, is whether this low adoption pattern also persists to other online safety practices, such as those for privacy and identity theft protection. Moreover, there's limited knowledge about what happens after consumers adopt advice, such as how often they abandon this advice and why. Next slide, please.

For our research questions, we scoped them with security, privacy, and identity theft as three key dimensions for online safety. First, which online privacy and security practices are fully adopted, partially adopted, or abandoned; second, what factors predict the level of adoption; and third, why are certain practices partially adopted or abandoned? Next slide, please.

We selected 30 expert recommended practices in all three domains from prior work. We included 12 security practices from Ion et al's 2015 study. We included 12 items from the US Census Representative Survey by the Pew Research Center for Privacy Practices. And we included six items from FTC's online resources for identity theft practices. Next slide, please.

Here I want to give an overview of practices we examined. Security practices include two-factor authentication, antivirus, cautious click and behavior, good password habits, and so forth. Privacy practices include a management of one's browser extensions and cookies, careful online disclosure, use of VPN and encryption, among others. Identity theft practices are a mix of

services provided by credit bureaus, commercial services, and manual tracking of credit reports and statements. Next slide, please.

As an overview of our method, we chose to conduct a survey and recorded 902 participants for the study on Prolific. All participants were US residents, since some of our examined practices are specific to the US context. Next slide, please.

The main survey questions were about 10 practices randomly selected from our list of 30 to not overwhelm participants. With 900 participants, this resulted in about 300 data points per practice. Participants could select if the practice was something they always did, did it with exceptions, did in the past but abandoned, consider doing, rejected, or were not aware of. Next slide, please.

At the end of the survey, we collected information about demographics, technical background, and prior negative experience. All participants' gender and income distributions are representative of the US population, but are skewed to younger and more educated people. Next slide, please.

Onto our findings. First, what practices were adopted or abandoned the most? Next slide, please. We found high adoption of security practices indicated by the deep blue in this graph. Interestingly, the top two adopted practices had to do with cautious clicking-- 95% for click links in emails and 93% were attachments in emails. Next slide, please.

We found that the most abandoned practices tend to be privacy related, though practices were not often abandoned, and the abandonment rate was below 20% for all practices. Looking at the light blue bars in this graph, practices with the highest abandonment rates were using anonymous systems like VPN, use fake identities for online activities, and clean web browser cookies periodically. Next slide, please.

By contrast, the adoption of identity theft protection practices were concerningly low. Looking at this large area of red and orange, most participants were either unaware of or rejected these practices. Credit freezes and fraud alerts, though strongly advocated by the FTC over the years, were among the top rejected practices, with more than 50% rejection rate. Next slide, please.

Onto answering our second question-- what were the factors that influence a practice being fully adopted, partially adopted, or not adopted? Next slide, please. For factors related to the practice, we confirmed that adoption levels of security practices were significantly higher than the other two domains.

Additionally, we divided practices into three subcategories-- manual practices that rely solely on user efforts, such as avoid clicking suspicious links; automated practices that after initiated require no user effort, such as running antivirus software; and finally, assisted practices that use tools but still require regular user interactions, such as two-factor authentication. What we found is that assisted practices were adopted significantly less than manual or automated practices. Next slide, please.

We also examined factors related to the user that influenced adoption. Experts had high levels of adoption than non-experts. And we further unpacked this difference and found that computer science and IT expertise more so than privacy and security expertise significantly impacted adoption rates. Additionally, being a previous victim of identity theft made someone more likely to adopt protection practices across all three domains. Our paper includes more details about other findings related to demographics. Next slide, please.

We analyzed participants' open-ended responses to understand why they partially adopted or abandoned certain practices. Next slide, please. Regarding reasons for partial adoption, the most common one is only adopting practices for certain sensitive sites, which is the case for private browsing. Another 10% of participants selected partial adoption, said the practice was inconvenient and difficult to use consistently-- for instance say, if I'm in the middle of doing something I won't be able to install this software update, or say that it's hard to keep track of unique passwords for different accounts. Next slide, please.

Regarding reasons for abandonment, 20% of participants who used but then abandoned a practice say they don't need the practice anymore, as it does not provide sufficient values to guarantee continuous usage. Like, I have used it but don't find it all that helpful for private browsing. Another 14% reported abandoning a practice when the perceived risk has diminished after a negative event. For example, I had a credit freeze due to suspected identity theft in 2012. But after some years, they decided to not use the freeze anymore. Next slide, please.

We discuss how our research has implications for how experts can provide online safety advice to consumers to increase adoption and reduce abandonment. Next slide, please. To bridge the gap that security practices were adopted much more than privacy and identity theft practices, it's important to show the synergy that exists between practices, especially in cases when multiple practices could add additional protection layers.

For instance, to combat phishing scams, avoiding clicking on the links is a common security tip. But this advice could be complemented by recommending users also actively monitoring their financial accounts as an identity theft protection tip, but also an important mitigation practice after one has fallen for phishing. Next slide, please.

Using an FTC's online article for identity theft self-protection as an example, this could be improved by giving more guidance as to which practices are most important to adopt and the connections between different practices and how they mutually benefit each other. For example, with measures for keeping your personal information secure online versus offline, we can illustrate how they work together and why it's important to do both. Moreover, it's important to identify the most effective and urgent actions to be prioritized so that consumers are not overburdened to take all actions at once. Next slide, please.

We can also leverage at-risk situations for communicating advice given the finding that experiencing identity theft drives the adoption of online safety practices. In case of a data breach, consumer-facing data breaches notices can be a possible value for education. Consumers reading these notices will be highly motivated to resolve the situation and mitigate future risks. And so,

resources that encourage and explain how to adopt protection practices will be most effective at that moment, though the advice must be actionable. Next slide, please.

We discuss how current tools for consumer online safety protection can be improved. Next slide, please. We found that usability issues prevented the full adoption of practices across all three domains. This echoes previous research in computer security about 2FA, password manager, software updates, and encryption, et cetera. Though these are security practices, in our study we also found evidence of usability issues with privacy and identity theft protection practices as well. Next slide, please.

This calls for more systematic research to better understand what these usability issues are and how to solve them. And another potential idea, more relevant to lawmakers, is to require usability testing for provided tools so that they are not made hard to use intentionally, which can reduce the burden on consumers. Next slide, please.

As examples for requiring usability testing, we can think about requiring readability testing in data breach notification laws to ensure that breach notifications are readable and reduce the chances of them being lengthy and full of jargon. We can also think about [INAUDIBLE] patterns in mandated privacy notices and controls to give consumers real autonomy in privacy and data choices. Next slide, please.

To summarize, for our study we studied the adoption and abandonment of various online safety practices. We find different patterns of adoption and abandonment between security privacy and identity theft protection practices. This implies the importance of expert advice to emphasize that synergy exists between practices and two, that more work is needed to improve the usability of privacy and identity theft tools in order to reduce user friction and encourage long-term adoption.

Feel free to refer to our paper for more details and reach out to me if you have any questions. Thank you.

JAMIE HINE: Excellent. Thanks so much, everyone. We really appreciated those presentations. Let's move into Q&A. Just a reminder, if you have any questions, feel free to send them through the privacycon@ftc.gov address, and we'll try and reach out to-- get to some of those.

So the first question I have actually is for Hana. I wanted to first ask you-- one of the conclusions that you reach in your paper is about notice and consent, which you rightfully mention is sort of a dominant approach here in the United States. But you suggest that consent mechanisms have failed to provide consumers meaningful privacy protections.

And so my question for you is whether your analysis justifies some type of an alternative approach. And the hard part of the question is, if there is one, what do you think that should be?

HANA HABIB: Yeah, I think going forward there still is a place for notice and consent. But it really needs to look a lot different from what it currently looks like now, which is typically you go to a website, we see a wall of text, and then you click a box that says, I agree, which doesn't necessarily translate to meaningful notice or meaningful consent, because people don't really

know what they're agreeing to. We can potentially replace that with interfaces that allow people to make their preferences known upfront. Like I mentioned in my presentation, there is a potential for having tools built into the web browser, for example, where you set your preferences there, and those preferences are automatically communicated to websites without the user having to do anything other than that initial step of setting those preferences to begin with.

And I think we should also consider what people should be consenting to. Is it specific usage of information? What inferences can be made based on the data that's collected? So I think that's a space that needs to be explored in more detail.

So in general, I don't think my work advocates for replacing notice and consent entirely, just maybe rethinking what that should look like in the future.

JAMIE HINE: So Ido, if I can actually ask you the same question-- I think that your research also suggests that consent mechanisms have failed to provide meaningful privacy protections. And I'm wondering if you agree with some of Hana's conclusions or you think differently about that.

IDO SIVAN-SEVILLA: I totally agree with Hana's approach. I think consent became a meaningless term in our digital society. Users do not really understand what they agree for. They don't have real alternative to choose from to get the service. Recent studies from Helen Nissenbaum and Christine Martin about what users actually think about information flows reveal that when users getting aware of what's happening, they would never consent to what's going on behind the scenes of our favorite websites and mobile apps.

And I think user awareness is critical to pivot around and change what's happening in this industry. And one way to increase awareness is to visualize what's happening. There is an add-on to Firefox from [INAUDIBLE] a commercial company, to actually visualize what's happening. How many third parties are approaching you dynamically. And it's starting to get a sense of what's actually happening when you go to your favorite websites.

So this is one step forward. Users need to be much more aware of what's happening. And a complementary part of that is to require more transparency from these companies. How do you actually use my data? How do you class information about me? You can identify me in different context of the web. But what do you do with this information?

That's what we call the server side analysis of things, which is going to be kind of a black box to understand how these companies are actually using our data. This is our data. Remember, we are the data subject, and we have no idea what's happening with this data. So user awareness one, more transparency on behalf of the industry second. These are first two steps to get us out of this disturbing path.

JAMIE HINE: So I just want to follow up. You mentioned a browser extension for Firefox. And Firefox is used by a relatively small portion of consumers in the country. So apart from sort of increasing the transparency about how the information is used, how do you think that we increase adoption for either creation of tools or tools that consumers can use to actually exercise

their choice once they understand how their data is being used and they conclude they want to exercise choice to control or limit that usage?

IDO SIVAN-SEVILLA: Yeah, that's a great question. First, we need to obligate service providers to provide alternatives for consumers. Firefox have done a very interesting step by preventing third-party cookies altogether. This is a great step for our privacy. But you see that the industry is now calling this the past cookie area and moving to other ways to identify us and create fingerprints for our browser habits and operating system characteristics to know that we are the same person as we go over the web.

So the industry will always find its sneaky ways to circumvent and go around. It's kind of a cat-and-mouse race for our privacy. So we need to make them more transparent about what they actually do to us. And then once we have this in place, it's for us consumers to decide what we actually want to do and weigh our options. But we have to have alternatives for the first place.

And unfortunately, we have no transparency, no alternatives. And the situation is not so encouraging.

JAMIE HINE: Great. I want to open it up. Yixin or Daphne, do either you have any response to that?

DAPHNE YAO: I agree.

JAMIE HINE: OK. Sounds good. So Daphne, I want you to pose the next question to you. One of the things that struck me so much in your findings was that scanners, at least in the PCI contexts, appears need some significant improvement. And it sounds more generally like some-- either off-the-shelf or even open-source scanners sometimes outperform ones that cost thousands of dollars, or for some things, like cross-site scripting or SQL injections for example, you may not really be able to find a reliable scanner to identify those vulnerabilities.

And I think about that in the context of the FTC and the work we do in the privacy division, where a number of the companies that are under order are required as part of those orders to engage in scanning and use tools to identify vulnerabilities as part of their assessments. So I guess the question to you is, are your findings more broadly applicable? Or are you just finding this to be a problem within sort of the PCI world?

DAPHNE YAO: Great question, Jamie. It's definitely more broadly applicable. Some of the products that we tested are packaged as web scanners. So they have no mentioning of PCI, but then they still fail in some of-- a lot of the application level tests.

And part of the struggle that we find is it just is so complicated. Because if you think about it, I would not-- now I will-- but a typical researcher would not say, OK, I have tenure to complete, I have a PhD thesis to advise. Let's choose to build a deployable grade the cross-site scripting detector.

No one in their right mind will do it, because the minute you submit the paper, you will immediately get rejected. The reviewer in most conference will say, oh, this is not novel. We know about this attack. We know there is some way of, you know, conceptually how to detect it. Why am I reading this?

And the community, the research community, needs to change. It is changing slowly. I'm managing some conferences that try to push in this direction, deployable and impactful security. You know, you consider it novelty, but then you also need to close the gap. There is this big gap between security theory and practice.

And then you need to reward researchers. Somehow, you want to encourage people's vending efforts, sacrifice their time, and at the risk of not getting tenure to meet this, reduce the gap and meet the needs. And this is just a huge, huge demand. And so a lot of the-- I think it's a widely-- it's not just the web scanner. Many, many other aspects of security also need that those kind of tools, open-source tools that will be able to push the standards of the industry.

If you think about the profitability, for-profit companies, the minute they put up a product, they will not list all the limitations, being against their interest. And so they will vaguely say, oh, you know, we cover this, cover that. And then so it's only-- if researchers don't do this, don't do the measurement, don't provide transparency, no one will.

And then security is something that there is no silver bullet, everyone knows, and there's no guarantee. And it's all, you know, the devil is in the details. And so you have to know what cases to cover, so what is the gap, what is missing, what is my attack surface. So that needs a lot of work.

JAMIE HINE: So I want to move over to Yixin quickly. But I want a quick follow-up, Daphne. I'm curious if you could very briefly talk about what the reaction from PCI was. Because if you've identified scanners, and you believe that there may not be commercially available scanners to find certain types of vulnerabilities, how does an organization that requires that type of compliance reconcile the fact that there may not be tools out there that reliably identify those vulnerabilities?

DAPHNE YAO: Yeah, great question. So the person that I had a long conversation with from the Security Council fully acknowledge our findings. And I do understand their struggle. So basically, they have two testbeds, they test the scanners. But then because the industry practice is a lot-- the level that we understand how to solve the problem is together collectively low, but then they have to certify some scanners.

And so they have to reduce their bar to a certain extent. And then PCI, they have built a very strong community. I really was very impressed that they help scanners to pass their tests. And so in that kind of thing, you know, they do have scanners. They said they kick out a lot of scanners out of their approval list. But then it's a problem that they have, they also struggled with, that if everyone fails the test, then this test is not very meaningful.

JAMIE HINE: Excellent. Thanks so much, Daphne. Yixin, I wanted to talk about-- your research touches on usability. And it includes recommendations, for example, practices and tools, to improve security, privacy, and identity theft. And so my question to you is, what do you think is driving this disconnection between users and interfaces?

Is it just poor interface design? Is it just developer laziness? Or could it be consumers? Are consumers just simply unwilling to take responsibility for their privacy and protection on the web?

YIXIN ZHOU: Thank you, Jamie. That's a great question. So I guess my immediate response will be, I don't believe it's the incompetence of all developers, designers, and engineers. I think we have people capable of doing this.

The issues I see are probably three-fold. First is the lack of understanding for usability issues. Like, in my work I see this is well covered for security practices, for some of the privacy practices. But I've yet to see like a comprehensive audits of major identity theft protection tools, even though our study has shown anecdotal examples from certain survey respondents. But we need a better understanding of what the issues are in order to solve them.

And then second is not to blame users, but we need to realize the fact that most consumers don't have comprehensive understanding of technology, have limited knowledge, literacy, and also time. So for consumers, we need better education, more targeted, effective education, to make them realize these are the available tools and how to use them, by giving very actionable guidance.

And then the third part, I think more for regulators, is to think about how to motivate companies to design usable tools. And things like what I mentioned in my presentation, of the audits or patterns throughout mandated privacy notices and controls that I think regulators are already working on this, this will be a very meaningful step to ensure companies are incentivized to solve their usability issues, not intentionally making them hard to use because that's for their own profits.

JAMIE HINE: Hana, would you also like to comment on this? I think some of your work touched on some of these issues.

HANA HABIB: Sure. So yeah, as Yixin mentioned before in your presentation, I think the need for user testing is there. Like, you can't really produce these tools and expect them to work great off the bat. Developers aren't their users, so unless they have the actual tool in front and interfaces in front of real people who are using these tools as they would in their normal lives, they really have little to go on in terms of what problems people might encounter and what may be difficult for people to understand.

And additionally-- and I wanted to make another point-- I don't think it's that people are incompetent in terms of-- or that they don't really care about their privacy and security. In fact, I think the opposite is true. And that's what the research overwhelmingly shows.

I think it's more that security and privacy typically aren't people's primary tasks. They're usually-- people are using websites, using applications to do something else, really, not really to come up with a strong password, for example, any thing like that. So typically, privacy or security might be in the way of them doing their primary task.

So rather than putting the burden on users to make sure their privacy and security is taken care of, it should really be on the part of companies to have better privacy and security practices. And I think that's where regulation can have a major role.

JAMIE HINE: OK. So there was one question from the audience. And that is to Hana. And we'll finish up with that. And the question is, did you notice any patterns in terms of the type of site and how difficult or easy it was to find opt-out information? So for example, were e-commerce sites more challenging to navigate versus gaming sites or popular news sites?

HANA HABIB: Yeah. I didn't get into this in the presentation, but we provide a little bit of details about this in our paper. So the way we sampled the websites was that we picked the really popular websites from the top 10,000 list as well as some less popular sites and sites that really probably most people haven't heard of. So we call them top, middle, and bottom sites in our paper.

And one positive note that we noticed is that, across the three different categories there really wasn't a difference in terms of the number of privacy choices being offered. But how and where they were offered seem to vary. So for top sites, for example, they generally had controls within the account settings as well as somewhere else in the website, like a privacy policy, or even like in About Ads page, dedicated About Ads page for websites that had targeted advertising, whereas, the middle and bottom web sites relied more heavily on the privacy policies to provide consumers these choices.

Additionally-- I guess that's the time.

JAMIE HINE: Please finish your thought, sorry.

HANA HABIB: OK. Yeah, so additionally, the way that these choices were provided in the case of targeted advertising opt-outs, for example, the more popular top websites tended to have their own implementations of these tools and a setting within that, a setting for that, whereas other types of websites relied more heavily on third-party opt-outs offered through like the Digital Advertising Alliance or the NAI.

JAMIE HINE: Excellent. On that note, I just want to thank all the panelists so much. This has just been an absolute pleasure to moderate. The research, fantastic. I invite everybody on the web-- please go to the Event page and check it out.

Everyone from this page is going to check out, and I have a few closing remarks in just a moment. Thanks again.

HANA HABIB: Bye, everyone.

JAMIE HINE: Thank you, all. [INAUDIBLE]

Excellent. Well, that brings us to the end of our fifth PrivacyCon. We are so thankful for having everyone here today. As Elisa mentioned earlier today and was obvious to everybody who was here, we moved things virtual. And I think it went really well.

There were a lot of changes that happened moving from a live event to a virtual event. And we just want to thank everybody for your indulgence today. We want to thank all of the panelists. Please know that they worked really, really hard to deal with technology and to make this such a great event.

Just a few thank you's again. I know that Elisa thanked a lot of people today, but I just want to thank all of the moderators. I want to especially thank Elisa, who worked so hard with me to help put this together.

I want to thank all of the support that we had. There are people like Leah Singleton who has gone out of her way to help make sure that all the slides were as perfect as they could be. Alex Iglesias helped with every technology issue that we couldn't figure out. Sheryl Thomas was on top of all of the Twitter today to help make sure that we sort of got the word out on social media. There are so many other folks behind the scenes at the FTC that had to work twice as hard to make this happen virtually. We want to thank all of you.

A few last remarks. We will have the video up on the Event web page on the ftc.gov. In a couple of days, you'll be able to go revisit and see all the presentations again. On the agenda, almost all the papers are linked so you can access all of those papers. If there are updated versions in the next few weeks or months, we'll update them accordingly.

And I just want to remind everybody that PrivacyCon is an event that we started several years ago to help create relationships with people that are doing amazing research. And that's not just about PrivacyCon. We have the privacycon@ftc.gov address. We also have research@ftc.gov.

We encourage anybody who's doing interesting work that you want to share with the FTC, please send it to us. Send it to us any time of the year, we're happy to take a look at it. In the next few months, we'll be announcing our sixth PrivacyCon, which will probably happen some point next year, same time, maybe online, maybe virtual again. We'll see in the next few months.

But we want to thank everybody for participating. And we hope to see you next year. Thanks again. Take care.