PrivacyCon 2019 session 1

ANDREA ARIAS: Here at the Federal Trade Commission. I'm happy to welcome you to our fourth— that's right. —fourth PrivacyCon today. We're happy to see so many of you here today. My name is Andi Arias, and I'm an attorney in the division of Privacy and Identity Protection here at the commission. My co-organizer for today's event is Jamie Hine, also from the division of Privacy and Identity Protection.

Before we get started with our substantive program I need to review a few administrative things. First of all, please, please, please silence your mobile phones and other electronic devices. If you must use them during the workshop, please be respectful of the speakers and your fellow audience members. Please be aware that if you leave the Constitution Center for any reason whatsoever, you will have to go back through security screening again. Please bear this in mind and plan ahead, especially if you're participating on a panel so we can do our best to remain on schedule today.

I've been asked to emphasize the next point, so please listen closely. Most of you received the lanyard today with a plastic FTC security badge. We reuse these for multiple events, so when you leave for the day please return your badge to security.

If an emergency occurs in the building that requires you to just leave the conference room, but not the entire building, please follow the instructions provided over the PA system. If an emergency occurs that requires you to evacuate the entire building, an alarm will sound. Everyone should leave the building in an orderly manner through the main 7th Street exit, 7th Street exit. After leaving the building, turn left, and proceed down 7th Street across E Street to the FTC emergency assembly area. There'll be people there with the right documentation showing you. Remain there until instructed to return to the building.

The restrooms are located in the hallway right outside this auditorium. The Plaza East cafeteria is located inside the building. So you can go without having to go through security again. It is open until 3:00 PM, and I've been told that's a hard stop. So please be mindful and plan ahead. Please remember, however, that no food or drink other than water is permitted in the auditorium.

We will be leaving time for audience questions during each of the panels today. For those of you in the room, if you'd like to ask a question during a panel, raise your hand and one of our colleagues will hand you a comment card to bring up to us. For those of you participating via webcast, hi, everyone out there. FTC staff will be live tweeting today's workshop at #privacycon19. So if you would like to ask a question via Twitter, please tweet your questions using @FTC and #privacycon19.

Please understand that we may not be able to get to all of the questions. Please be advised that today's event may be photographed and it is being webcast and recorded. By participating in this event, you are agreeing that your image and anything you say or submit may be posted indefinitely at ftc.gov or on one of the Commissioners' publicly available social media sites.

We're happy to welcome those watching via the webcast. We will make the webcast and all of the workshop materials available online to create a lasting record for everyone interested in these issues. Lastly, I want to say thank you to our researchers and panelists for taking part today. We’re grateful for your time and work in the privacy and security area.

Aside from the folks that you’ll be seeing on stage today, this program would not be possible without the great work done by many of our FTC colleagues. Please indulge me in letting them thank them for a few minutes. We want to thank our colleagues that assisted us in reviewing all of the research submissions including Dan Salsburg, Lerone Banks, Tina Young, Phoebe Rouge, James Thomas, Yan Lau, and Ryan Meem; and those moderating panels today, including Marc Eichorn, James Thomas, and Yan Lau.

Finally, this conference would not be possible without the help of Crystal Peters, Arisa Henderson, and Bruce Jennings. Alongside our paralegal support today from Soojin Jeong, Ryan Sullivan, Ashley Knott, Patrick Curtain, Tyria Bunche, Courtney Butterworth, and Emily Liu; and support from Leslie Fair and June Chang from our Division of Consumer and Business Education and Nicole Jones from our Office of Public Affairs. Thank you all. Now it is my distinct honor and pleasure to welcome the Chairman of the Federal Trade Commission, Joseph Simons.

[APPLAUSE]

JOSEPH SIMONS: Well, good morning, everyone, and welcome to PrivacyCon 2019. During my first stint at the FTC— I’m in my third now. My first stint was back in the 1980s, way back in the 1980s. Personal computers were just being introduced into offices and homes. No one imagined that we would be soon carrying them in our pockets, speaking commands to them, or using other devices to track our fitness regimes, unlock our doors, and control our thermostats.

I can remember back when I was a young associate in a law firm just out of law school, Compaq came out with a computer that they referred to as portable. It weighed about 40 pounds. What a difference. Few of us then envision the advances in technology we would experience in our lifetimes and the effect they would have on our everyday lives. Even fewer of us had the foresight to recognize the commodity unifying these technological advances – our data.

When consumers engage digitally, companies collect information about their choices, experiences, and individual characteristics. Every day, companies make countless decisions based on our likes and our dislikes, our relationships and conversations, our transactions and our purchases. They carefully assemble, synthesize, trade, and sell these small bits of data, providing insights into market wide tastes and emerging trends, allowing for their prediction of individualized preferences.

No doubt this vast amalgamation of data has allowed for great technological advances, but it also comes with risk. News stories highlight troubling privacy and data security practices on a regular basis, whether it’s allegations of using facial recognition technologies and images without users consent, breaches that expose health data, or sharing genetic data beyond consumers’ expectations.

These types of privacy and data security failures don’t just generate headlines. They can cause real harms, including fraudulent charges on credit cards, safety risks, reputational injury, and unwarranted intrusions into people’s homes and the intimate details of their lives.

In part, to examine these types of incidents and the injuries associated with them, we hosted our first PrivacyCon back in 2016. Since then, PrivacyCon has been an annual event that has enabled us to advance our consumer protection mission in multiple ways. It has allowed the FTC to stay up to date with emerging technologies. It has helped us identify potential areas for enforcement and to fashion our remedies in better ways, and it has highlighted areas in which we can provide additional business and consumer education.

This is my first PrivacyCon as Chairman. And as you undertake your discussions today, I thought it would be useful to hear about some of the FTCs current priorities on privacy and data security. First and foremost is vigorous enforcement. Where we have statutory authority, we use it to the full extent. In the past year we bought privacy and security cases under the laws we enforce. And in the limited areas where we have civil penalty authority, we have used it aggressively and we expect more in the future.

In February, we announced our highest penalty in the children's privacy case against TikTok, which is a popular video social networking app. Last fall we obtained a $3 million civil penalty under the Fair Credit Reporting Act against the company whose automated decision making tool provided inaccurate data to property managers resulting in denial of housing.

Second, I've been very focused on improving our non-monetary remedies in privacy and security cases in order to provide better deterrence. As part of our hearings on competition and consumer protection in the 21st century, we hosted a data security hearing which included a panel specifically focused on the FTCs data security enforcement.

Partly in response to feedback that we received during the panel, we have incorporated new provisions into our data security orders. For example, in three recent cases we required that senior officers provide annual certifications of order compliance to the Commission, thus improving individual accountability. While we continue to require that companies implement a comprehensive process based data security program, in our most recent case we also included specific requirements that the company conduct yearly employee training, monitor its systems for data security incidents, and implement access controls.

We've also made significant changes to improve the accountability of the third party assessors that review the company's data security programs, requiring that the assessor look "under the hood," rather than relying on the company's assertions. And also, we've created greater oversight of the FTC for the assessor allowing us to hire and fire them. Third, we continue to use all of our non-enforcement tools at our disposal, to further our privacy and data security mission.

For example, we have proposed amendments to the safeguards rule, to add more detailed requirements. We have used our authority under 6(b) of the FTC Act to request information from several ISPs to examine how broadband companies collect, retain, use, and disclose information about consumers. Finally, we've engaged in advocacy. Recognizing the limitations of our primary legal enforcement tool, Section 5 of the FTC Act, we have urged Congress to enact privacy and data security legislation, enforceable by the FTC and which grants the FTC civil penalty authority, targeted rulemaking under the APA, and jurisdiction over non-profits and common carriers, which brings us to today.

We are using yet another tool at our disposal, PrivacyCon, to continue promoting privacy and data security. Today's program has four sessions that will address a variety of important topics. Our first panel, Privacy Policies, Disclosure, and Permissions, will explore privacy policies, how data collection aligns with those practices, and the GDRs impact on both privacy on the web and on apps.

Our second panel entitled Consumer Preferences, Expectations and Behaviors, will examine consumer attitudes towards digital privacy and take a deeper dive into the internet of things, devices, smart homes, and COPPA. Our third panel, Tracking and Online Advertising, will consider the commercial impact of tracking technologies, free versus paid apps, and the GDPRs effect on e-commerce. Finally, our fourth panel, Vulnerabilities, Leaks, and Breach Notifications, will consider the data security aspects of apps and the effectiveness of breach notifications.

I know we are all excited for these presentations to begin. But before we do, I want to thank everyone who's made this event possible today. First, I want to thank the 19 presenters here today and their dozens of co-authors who submitted research for today's event.

Thank you to Andi Arias and Jamie Hine for leading the planning of this PrivacyCon. And I also want to thank the many other FTC colleagues from the Division Privacy and Identity Protection, the Bureau of Economics, the Division of Business and Consumer Education, the Office of Public Affairs, and the Office of the Executive Director, who have worked together to produce this wonderful event.

Finally, thank you to everyone who is attending in person or watching online. We very much appreciate the opportunity to engage the public on these important research endeavors. And I hope you will all enjoy this PrivacyCon. Thank you for coming.

[APPLAUSE]

JAMIE HINE: Good morning. My name is Jamie Hine. I'm an attorney in the Division of Privacy and Identity Protection. My co-moderator this morning is Marc Eichorn. He's an assistant director in the Division of Privacy and Identity Protection. Our lead off session this morning is on Privacy Policies, Disclosures and Permissions. This morning, you're going to hear from five researchers who will each have 10 minutes to provide a summary of their work. And we'll follow those presentations with a 20 minute discussion session.

While the questions will follow, please start sending any questions you have now while you're watching the presentations. Anyone here in the audience can use the comment card, simply raise your hand, someone will come by and collect them, bring them up to us. And for those of you who are watching by the Web, please tweet us. Use @FTC and the #privacycon19. We'll receive some of those and we'll try to get to as many of the questions from the audience as possible.

So let me briefly introduce our presenters for this first panel. The full biographies for each of the participants and their funding disclosures are available on our websites and in the agendas that were available this morning. So first, to my left is Yan Shvartzshnaider of NYU and Princeton University; to Yan's left is Kassem Fawaz from the University of Wisconsin-Madison; to Kassem's left is Justin Brookman from Consumer Reports; to Justin's left is Christine Utz from Ruhr University Bochum in Germany; and finally, though certainly not least, we have Jonathan Schubauer of Indiana University. Yan, would you please start us off with your presentation, Privacy Policies Through the Lens of Contextual Integrity?

YAN SHVARTZSHNAIDER: Thanks. All right. Thank you, everyone. Really great to be here. That’s me. All right. So really glad to be here, and today I’m going to discuss a work on analyzing privacy policy using a theory of Contextual Integrity. This work is a collaboration between New York University, Princeton Center for Information Technology Policy, and Cornell Tech.

So as the chairman mentioned today, we’re all using apps and services and have all kinds of gadgets and smart home devices in our homes that essentially collect and share information. And the question we like to answer in our work and I presume throughout the day today, “Do these services actually respect our privacy?” That’s not an easy question to answer. But we thought a good place to start would be a privacy policy where a company is supposed to disclose some of the practices that they use. And, we would like to know whether their resulting information flows from those practices essentially conform to our privacy expectation.

To do that, we’re essentially using the theory of Contextual Integrity. And I really hope most of you in this room know about the theory, but even not, I really recommend reading the book Privacy in Context by Helen Nissenbaum where she outlined, basically defined privacy as not a secrecy or control of information, but rather as an appropriate information flow. And appropriateness is defined by the contextual information norm. And essentially what it means, when we actually use, and according to the theory when the users engage the service, they come with privacy expectations in mind. And the theory provides you with a framework of five parameters that essentially allows you to capture those information flows and see whether they’re aligned with those privacy expectations.

Those five parameters are: the sender of the information; there is the information type, which is the attribute; the recipient of that information; the subject about whom this information is, and what; and the transmission principle, which is the constraint on that information flow. And what is important is the values for all of those five parameters matter, all of them. In order to make a valid assessment of the privacy implications, you have to specify those values.

Let’s see, for example, that if in this information flow I am the sender of my medical data. So I’m the subject, I’m sending it to my doctor, and obviously with my permission if we change the recipient to, let’s say, a colleague, suddenly my medical information is going to my colleague, which might not really align with my expectation in the medical context.

So all of those parameters matters. in our work and through theory, we have to look for the senders, we look for recipients of information, the subjects which are essentially overall actor’s category. There’s the information type that’s being transferred, as well as transmission principle, which is the reasons and purpose and conditions for why this is actually happening.

So we look for all of those, and this is really the heart of our methodology. We’re essentially using the Contextual Integrity framework to annotate privacy policies with those parameters. We look for the senders that’s in this snippet where we look for the attributes test and the recipient’s transmission principles and so forth if they actually outline.

Once we annotate those, we essentially have a way to detect policy ambiguities. Now, today I’m only going to mention two of our analyses, but really I refer to the paper to see the rest of it and things that you can essentially do with those once the policies are annotated. So one thing we can identify, statements that are essentially missing one of those parameters. They are essentially omitting some of that context information that are current to the theory is actually required. We call them incomplete information flows. On the other hand, there is statements that essentially will come with hidden complexity because they introduce multiple instances of those parameters. And we call this phenomena CI parameter bloating, which I'm going to talk briefly next.

For our case study we chose Facebook because it provided us with a unique opportunity. When the Cambridge Analytica incident happened, Facebook came out and said we're going to update our privacy policy to inform the user about our practices and provide more information to the users about what's going on. So what we did, we essentially annotated the previous [policy], before the updated privacy policy, and then we annotated the updated version.

And our analysis indeed confirmed that Facebook introduced across the board. First of all, as you see on the figure, the light bars as indicated updated policy across the board, there is more information about more information flows, there's more information by the type of information being sent, the senders, the recipients, and transmission principles. However, spoiler alert here, more information doesn't really necessarily mean that there is more clarity.

And the reason for this is, so once we annotated the policies we see that in previous privacy policy we find 47% of privacy statements, they're essentially incomplete. They're missing one or more of those required CI parameters. The updated privacy policy, although it includes more information flows as Facebook stated, but also that number doesn't decrease. In fact that percentage has actually increased to 55%. Not specifying those parameters essentially creating gaps. So somebody who reads the privacy policy will substitute those mentally with their mental privacy model, which basically leads to them being uninformed on actual practices of the company.

As I mentioned on the other hand, there is the CI parameter bloating phenomena. As in this example, this privacy statement introduces three unique senders, five information types, and six transmission principles, essentially conditions under which the information is being transferred. So what this actually means, that essentially the statement provides you with a permutation of all of those parameters. So resulting flows are not just one flow, but a combination of all those parameters, which in some cases can lead to quite a bit of flows that essentially the user, or somebody who is reading the privacy policy, has to comprehend.

And through our analysis we see that on average there is over 10 information flows per those statements, and sometimes the outliers go to close to 100, and the real extreme is over 100. Information flows generated basically from those overloaded privacy statements. So this hidden complexity really, if the user wants to see that those conform to privacy expectations, they actually have to check each of those information flow to see that. And we cannot expect that to happen realistically.

So we feel that we need more to scale this approach. So we tried crowdsourcing in our work. We essentially created a CI annotation task and deployed it at Amazon Turk. We have initially really promising results — high precision, and slightly lower recall. And this basically from our observations shows that essentially Turkers were able to annotate privacy policies correctly. When they were actually annotating, they were doing it correctly. They weren't actually always annotating.

So in our paper we discussed this and some of the reasons really fundamental to the crowdsourcing approach and the platform itself, but we also provide some ways to mitigate that in the future as future work. And, essentially our ultimate goal is to create a corpus of annotated privacy policies to contribute to the community and essentially allow us to detect some of those ambiguities and essentially look for trends across and within industries.

Now, the main takeaway here— and I would like to spend a few moments here. —is that really good privacy policy should really conform to privacy expectations. Now, I know after all the talk I said users' privacy expectation. That doesn't necessarily mean that's the only stakeholder here. So if the regulators, the lawyers, whoever reads the privacy policies, should be able to see what they say their practices to describe. The resulting information flow should really conform to whatever norms are required.

Now, that means that omitting relevant contextual information leads to ambiguous and misleading policies in statements. So the idea here is you don't want consumers to guess, or somebody else who's reading the stakeholders to guess, or you actually have to be explicit what information flows actually going to result from your practices. The other side of it is actually that CI parameter bloating leads to hidden complexity that generates those hidden information flows that consumers sometimes be beyond their [INAUDIBLE] efforts and to parse it and actually see that's conform to them. And we really don't want that if you like to be transparent about that.

All right. So I'm really happy to be here. I am right on time and discuss it throughout the panel. I would like also take this opportunity to also invite you to the Contextual Integrity Symposium, which is going to be in Berkeley in August where we discuss those issues and others and see how context integrity can help. Thank you.

[APPLAUSE]

MARC EICHORN: Next you'll hear from Kassem Fawaz, describing Polisis, automated analysis and presentation of privacy policies using deep learning.

KASSEM FAWAZ: Thank you. Good morning, everyone. This work is a collaboration between me, my colleague, Hamza, who couldn't make it because of Visa issues and for Florian sitting over there. So Yan made my job easier by motivating the problem. So privacy policies are these main documents that service providers use for privacy notice. They inform you how your data's being collected, processed, shared, and for which purposes.

Can I have like quick show of hands, who reads the privacy policy of each website they interact with? Yeah, who needs them? Not for research purposes.

[LAUGHTER]

OK, one person. And the PrivacyCon. Cool. And there's a reason, right. Privacy policies are long and there was this study in 2008 that if you would like to interact with the privacy policy of each website you visit, it's going to take you 200 hours per year. And that was in 2008. Now we have mobile phones and smart devices, so it's much of a bigger problem.

When we did these slides I didn't expect that I'm going to be presenting at FTC, so I kept the lawyer thing. I'm sorry. So researchers have identified that this is a real problem and they have come up with interfaces and user interfaces to present privacy policies in a more usable manner.

One of those is the nutrition table approach from CMU in 2009, and the main approach here is like let us standardize a privacy policy. Instead of having it this natural language document, let’s present it as a form of information being collected versus ways the information is being used, and then in the cells you have the opt in and opt out options. This is a great approach. We’re already used to nutrition labels.

The main problem is that it requires service providers to standardize their policies, and they did not. Another approach is from TLSDR.org, where they said, well, let’s rely on crowdsourcing for annotating these privacy policies. And they were able to annotate like 100 something privacy policies, but the main problem is that still requires manual effort and those tags are not standardized. So, you can’t create them automatically at the scale if you’d like to do some comparative study.

So the main problem is that manual work on annotating privacy policy doesn’t scale with the breadth of privacy policies and how privacy policies change over time. And besides, the interfaces failed to cope with new and emerging technologies. So we have these UI limited devices, voice-only devices, by which these UIs are not convenient to put like a nutrition label table kind of thing or a list of privacy practices. So you need some way to interact with these privacy policies that takes into account the UI limitations.

And I’m going to be describing later what I mean by unstructured queries, but what we envision that for this kind of devices is that you can have some sort of a privacy dialog through unstructured queries. And then there’s a question on querying privacy policies at this scale. Like in the previous paper that was just presented, the problem was, OK, let’s extract the CI elements from privacy policies. Can we do that at scale and extract it automatically? One other example is: can we test GDPR compliance of privacy policies at scale? And I’m going to show an example how can we support this?

So the main takeaways is that, manual effort doesn’t scale, existing interfaces, whether they’re APIs or UIs don’t cope with emerging and new needs. So we is something else, and this something else is what we call Polisis over here, which is an automated privacy policy analysis framework. The main takeaway is, if we can automate the analysis of privacy policy, then we can create them at a scale. Regular users can post questions, regulators and researchers can query them and perform competitive studies for compliance or for research purposes.

The main approach that we take in Polisis is we need to make a privacy policy machine readable. So given this natural language text, we process it, we assign it to a bunch of privacy tags that we can query. And I’m going to show a couple of examples. So we have three layers. So this is the most technical side of the talk. We have three layers. We have an application layer, we have a data layer, and a machine learning layer.

In the application layer, we take an input as a privacy policy link and the user query, and this is on a previously unseen privacy policy. We segment it, basically partition up to set of paragraphs, and then we pass these paragraphs each into a set of neural networks. Each neural network will assign the paragraph a privacy tag, which could be like the privacy category is first part of data collection, the data type of health information.

Now, instead of representing this segment or this paragraph by a bunch of forwards, it's represented by a set of standard privacy tags that are derived from a taxonomy that has been proposed by the Usable Privacy Project at CMU. And we uses their data set of PP150 to train these classifiers. We do something similar for the queries which you can look at as like a user question— I'm going to show an example. —and then it's up to the application developer to get these annotated segments, their queries, and try to map them and get with answers. So this is the high level overview of how this framework has been designed and implemented.

So I'm going to show two examples. One of them is unstructured query, the other is structured queries. So for structured query we show an example of automated privacy icon assignments. So there was this project, disconnect icons by which you can present the privacy practices of some companies by icons and then you can assign colors. And I'm going to show an example. And that's required manual effort, which is probably that reason you can't see those anymore online.

So this is an example of a structured query that's enabled by Polisis. So if you can't read the text, I'm going to be explain again. So there's an icon for expected collection and it shows whether other companies, which can be done as a query in Polisis— like get me all the segments that describe third party collection. And then it's about ad providers analytics, so get me all the segments that describe which have the purpose of advertising or analytics.

And there's the tracking. So what's the purpose of the data collection? Is it to track websites or get from all the segments applying another filter that has the tag of tracking or collecting on the website? That icon has a value of yellow if there's an option of opt out. So of all of these filtered set of segments, do they have opt out options? And if they do, we'll return a yellow icon. So this is an example of how you can query a policy.

Another example of a query could be what Yan just described, the CI elements. You can post those as queries and then you can create a policy for them. An example of unstructured query is a user asking a question. And there's a video here that should play. It's a 17 seconds video that should play that shows an example of how a user can ask a question about the privacy policy. And the question is, "Do you share my data third parties?" And this is the excerpt from the policy of khanacademy.org, that shows the answer to this question. And then you have a simplified button that just summarizes this privacy segment into more readable text.

So this chatbot we call Pribot by the way. So this Polisis framework, the chatbot, we have online on pribot.org. And since we had them online a year, year and a half, we have had more than 45,000 users of the app, more than 100,000 minutes on our apps, and we have already analyzed more than 20,000 websites. And those are automatically analyzed websites, meaning that no human effort involved beside the designing and labeling the data and training the classifiers. And refresh these policies every once in a while because they might change. So we keep them up to date.

So the main takeaways are privacy policies are the main medium of privacy notice, manual effort doesn't scale. So we propose Polisis, which is a unified framework for creating privacy policy. Hopefully it's going to assist users, regulators, and researchers with their needs. We had two applications, structured and unstructured queries.

Another example of structured query is like our follow up work on comparing the privacy policy landscape before and after the GDPR. We post GDPR compliance as queries in

Polisis, and then we analyze how privacy Polisis have changed. You can see it on archive. And then we had some track on our applications for some media articles. And if you're interested in visiting our website, so it's pribot.org over there. It has more applications on Polisis, like visualization, privacy policies, and so on. Thank you.

[APPLAUSE]

MARC EICHORN: We'll now hear from Justin Brookman who will be talking about the privacy practices of large online data platforms.

JUSTIN BROOKMAN: Cool. Thank you. So, yeah, I'm presenting here on behalf of Consumer Reports. It's just my name up there, but there are a lot of other people that have been involved. Katie McInnis who will be on the next panel, Bobby Richter who is now on his honeymoon in France. So good for him.

So our project was to try to take a holistic look at the comprehensive practices of a handful of large Internet companies just to try to see what is externally documentable about what they do? Caveat, this has not been published yet. We'll be putting it out later in the summer. So it's not currently available.

Briefly, I want to give like just one or two slides of context about the rationale behind this. This is part of Consumer Reports development of a comprehensive privacy testing regimen. Fortunately we've been able to present at the last couple of PrivacyCons about the development of this program. The short version is, Consumer Reports is really good at testing stuff over time, recognizing a lot of connected products that other attributes like privacy and security that folks might want to know more about, that there's not a lot of clear, digestible information made available about them today. Privacy, security, things like repairability.

And so part of the ongoing mission of the organization is to develop a repeatable, scalable, consistent, reliable tests for these attributes. And the good news is we're starting to actually do it. This is the first set of scores we released around peer to peer payment apps in the winter of this year. We've had more sense than we have— hopefully another batch coming out next month.

In this case, Apple Pay performed the best. They're structured to minimize data collection. There are strong, documented prohibitions on secondary usage. They seem to have, so far as we could tell, a strong security protocol. There's more clustering on the security side, the data privacy side. Apple performed by far the strongest on the test.

And so stepping back, we wanted to try to look at it more broadly a bunch of the largest Internet platforms and try to see if it possible even to develop a full story about the data behaviors of these companies, not even for ratings. I think we'll probably tend to try to rate them more on a product by product basis just to try to get a coherent sense of what is actually happening.

And so the companies we looked at are probably not surprising. The big five, a handful of Internet service providers, other companies that have a large view into behaviors like Twitter, Alibaba. I think we used Wikimedia there at the bottom more as a control to compare against maybe some of the larger companies.

In this poorly rendered circle on the right, this is actually documenting which they are. So the blue swath in the lower right, that's all Google. Google has 200 or so different products. And so can we get a sense of how data collected across those is used in different contexts?

The things we looked at: what data is collected; how it's used across these platforms; how and with whom it's shared, and where discernible for what purposes; data retention, when data is deleted by default— It will probably not be surprising to hear there was not a lot of information about that. Companies tend to reserve the right to maintain data as long it remains interesting to them; and control, what controls are given to users, kind of not by default.

So if someone makes an affirmative choice— and this is actually one thing that we're actually probably changing about our ratings after we gone through a few iterations, that we're looking probably more closely at default minimization practices and less at just things like mere transparency or the availability of an opt out control as far as assessing what conforms with reasonable expectations and what might be good or bad on privacy. But controls are still important and something we wanted to document. And given where personal preferences are divergent around personal information, are still very important.

So the things that we looked at. I mean we looked at the things I think are expected: privacy policies, terms of use, FAQs, marketing materials, which actually contain a lot of interesting information that's maybe a little bit more aggressively presented than maybe privacy policies; secondary sources actually— we looked at a ton more than the number there. I think the 86 is from new facts we got that we hadn't gotten otherwise. And then we looked at a lot of papers, actually a lot of PrivacyCon papers, as far as trying to dig in and look under the hood and document what was going on that was not otherwise available.

One of the thing we were most interested in looking for is inferences with the data usage. Do companies document how they tie this data together about you to make inferences about attributes? Again, there's not a lot of clear documented information about this. Actually in some ways the PrivacyCon papers were more helpful in identifying potential inferences. There was a Princeton paper here a couple of years ago about how IoT manufacturers could infer a fair amount of data from inferences about you just from how and when devices are used in the home.

Most of these companies make clear that it can be used for advertising. Not a lot of information on exactly how, and actually I think there's probably less information available than there used to be on this. I know maybe five years ago a bunch of companies made portals available. You could go and see what ad categories you're put in. In our review there's actually fewer of those than there used to be a couple of years ago. And the ones that we did find tended to have less information about that. We did not look to see what was available in Europe after GDPR. This could potentially change in America after CCPA, but not clear that would apply to inferred data.

Policy constraints is maybe one of those important things we looked at. How are companies self limiting what they do with data? All these companies can collect a tremendous amount of information about you, and then once they have it they could use, share, retain, without being actually testable. So with few legal limitations on what companies can do and maybe fewer technical limitations, self-imposed policy restraints are maybe the biggest actual constraint on what companies do.

Section 5 is kind of a blessing and a curse here. One is actually reliable. We can believe it, at least somewhat, when companies make these statements because they're enforceable under law. Because there's liability attaching, companies are actually more scared of making these sorts of policy documentation. And so actually we found that in our testing of products, we've actually been relying on these sort of statements a lot in assessing what's good and bad.

In the Apple example, Apple Pay scored well in large part because there were documented prohbitions on a lot of secondary uses. We've also looked at IP cameras that store data in the cloud. Again, not testable there, but the companies that put in place affirmative statements said they weren't going to do certain things with data that tended to score better.

We also did some collection measurement as part of this study. The place we looked specifically was on mobile apps that were put out by these companies. Mobile apps testing tend to be one of the more standardized things that are out there. Again, something that then presented at PrivacyCon a fair amount in the past. We worked with AppCensus, which is spun out of Berkeley's work. Looked at 400 apps from these 15 different Android apps from these 15 different companies.

These are from some of the largest companies, so you'd think that their behavior might be more constrained than some of the smaller players, but a lot of the kind of questionable behavior that has been documented by some other papers, we saw here a lot of questionable sharing of persistent IDs instead of advertising identifiers with third parties that seemed likely to be for advertising. In many cases, the identifiers were obfuscated, hashed, tried to obscure that it was happening. And at least 30 cases we saw serial numbers of phones being obfuscated and shared with cross site and measurement platforms.

Also they had a fair amount of sharing with each other. A lot of these companies also offered third party SDKs that people plug in. So we would see the Amazons and Alibabas sharing with the Googles and Facebooks and Verizons and Microsofts of the world. But we mostly relied on documentation.

And this is actually quite overwhelming. And this is not even just what is reserved because companies basically would reserved the right to collect anything in many cases. In one company what was documented that they say they collect, and each of these it can be emails, written, websites visited; it can be incredibly illuminating and a lot of inferences to be drawn from it. Overwhelming, but also maybe somewhat unsatisfactory in a lot of ways. This list is over inclusive perhaps, just reserve the right to look at it, and not a lot of clarity on secondary uses.

One quick takeaway, and I will go slightly over 30 seconds, the need for maybe more mandated transparency. People have privacy policies, not a lot of digestible information in them. Query who the right audience is for these, not perhaps consumers, but testers. And I think some of the tools that some of the people were working on to maybe automate these and create scores based on them over time I think could be useful.

Quick themes. A lot of limitations on sharing. Well, this is like the Facebook, we don't share your data, but not a lot on collection, or if someone use, but only on really more marginal cases. These are some of the preliminary themes that we saw. I look forward to sharing the paper later in the summer and talking more about the conclusions. Thank you.

[APPLAUSE]

MARC EICHORN: Next, Christine Utz will be talking about her study on the GDPRs impact on web privacy.

CHRISTINE UTZ: Yeah. Thank you so much for having me here. I'm really excited to be here and present our work about changes we observed on websites around the GDPR enforcement date. I guess everyone in here remembers the time in the month before May 25th, 2018, when everybody got tons of emails that web services are going to change their privacy policies, and they asked people to agree to that. And, there were lots of news headlines online like the ones shown here, and people were really unsure what to do.

And the new rules actually implemented by GDPR mainly are about transparency. Which means that if you have a web service that processes personal data, it basically means that they now need to provide a privacy policy. And GDPR also introduced new legal basis for the collection of personal data. And one of them's consent. And if a company wanted to base their data collection practices on consent, it is now expected to provide some information about that and ask for people's consent. So, basically this means we expect this company to display a cookie consent notice on their website.

GDPR also implements principles such as privacy by default and privacy by design. And the rules apply to any company processing the personal data of people in the European Union and not just companies based there. So with this in mind, we ask ourselves the following research questions. How did websites react to GDPR? Do we see increased privacy protection due to this privacy by default paradigm? Do we see more transparency on websites, and how do they implement these new consent requirements?

So to answer these questions, we collected a set of popular websites in the European Union by compiling a set of the 500 most popular websites in each member state of the European Union according to the Alexa ranking. And due to overlap between the different lists and changes over time, we ended up having a data set of more than 6,750 websites.

We used an automated browser set up to visit these sites once a month. And we did monthly scans from January 2018 to October 2018 and more frequent ones in May because we expected more changes to occur around the GDPR enforcement date. So when our browser visited one of the websites, it tried to identify the term for privacy policy in the 24 different languages of the EU. And once it had found such a document, we tried to download it. And the browser also looked for, we also searched for cookies and trackers on the sites, and we also took a screenshot of each website.

In addition, we collected all the versions of privacy policies from the Internet Archive. So in total we ended up having more than 100,000 privacy policies in 23 languages. For like 50% of websites we couldn't automatically find a privacy policy, so we had to do lots of manual annotation as well. And we also inspected all of the websites if they displayed a cookie consent notice.

So we found that the prevalence of privacy policies on websites has increased by about, all in all, 5% between January 2018 and our post GDPR scan. And there's really differences based on the individual countries. And I'd like to refer to our GitHub repository for a more detailed analysis of that. The link will be on the last slide.

We saw that the average text length of privacy policies increased by almost 50% between March 2016 and May 2018. And looking at the rate of change, we saw that 50% of all sites changed their policy between April and May 2018 alone. In all of 2018, we saw changes in 3/4 of websites compared to about 40% in the two previous years.

Regarding the content, we saw more privacy policies display some information required by GDPR such as the names and email addresses of data protection officers. And, we also did a search for GDPR related terminology in all of the different languages and we found an increase in terms associated with the new rights instituted by GDPR such as data portability or erasure.

Regarding the actual data processing that's taking place on the websites, we couldn't identify significant changes and we looked at HTTPS adoption to figure out if this data privacy by default mechanism was implemented. We saw a small increase, but this is in line with the general increase in HTTPS adoption because browsers have started to display warnings if websites are not secure.

The most visible change we saw was this increase in consent notices shown on websites. Overall this has increased by about 16% between January and our post GDPR scan. It's really popular that web developers use external libraries to implement consent notices. We identified about 30 different implementations and we did a technical analysis of these if they're actually capable of blocking cookies before consent has been given. And I'd like to refer to our paper for our detailed analysis.

We saw that there are really different types of interaction models in all the notices we saw. Here's a really simple one with just an OK button. And the simplest version probably is this one. It just doesn't give you an option. It just says yeah, the site uses cookies, and by continuing to use our site you agree to this use of cookies.

This is a problem because according to European Data Protection authorities, valid consent requires a clear affirmative act that needs to be freely given, purpose specific, informed, and unambiguous. And this implies that we need some active choice, and just continuing to use a site cannot really be considered active. We looked at how those different interaction models are distributed over the websites, and we found that 80% of websites offer you at most one choice, which is to accept the cookies. And even those banners that offer you the option to decline, for those we saw that often tracking would start before the user has clicked something.

This mechanism is a bit more fine grained. It allows you to select or deselect different categories of cookies individually, and this is in line with the GDPRs purpose based approach for consent. But here the problem is that often those more fine grain controls are hidden under some links such as settings and they are not immediately visible, so it might not be transparent to users what they consent to, and also the categories are often pre-ticked which contradicts the privacy by design paradigm.

Meanwhile, the online advertising industry has been pushing their IAP Europe transparency and consent framework. And in its most fine grained interface, it provides you with a list of up to 400 third party services users can allow or disallow individually, but no one's ever going to read through the whole list and make an informed choice for each and every single of these third parties. Also, the website may use third party services that are not listed here because they don’t participate in the framework and those consent banners don’t mention those. So these provide both too much and too little information at the same time.

So the conclusion we have arrived at after all of this. So we saw lots of changes on websites around the GDPR enforcement data. Most of them are related to transparency, like additions to privacy policies, but there are few changes in the actual data processing that’s taking place. The results are really different by country and somehow these new transparency requirements are at odds with the GDPRs overall goal to make this more transparent and make privacy policies easier to understand.

For the consent notices, there appears to be a lot of confusion about what’s considered sufficient or GDPR compliant and we think there’s a lack of guidelines from data protection authorities about that here. So here’s the link to our GitHub with the additional data. Thank you, and I look forward to questions later.

[APPLAUSE]

MARC EICHORN: Last, there’s Jonathan Schubauer from IU. Going to speaking on changes and app permissions over time and across jurisdictions.

JONATHAN SCHUBAUER: Well everyone, good morning. So today, I’m going to be talking about application permissions, what may influence them in practice, and how they have changed over time since recent privacy reforms. But before I begin, some of you may be wondering what is an app permission? You can think of an app permission as a form of privileged access on your phone. They govern what resources an app can access and utilize on your mobile device. There are many different types of permissions in practice today, but for the sake of time, I’ll be discussing with you two types of permissions overviewed in our study.

The first type was classified as a normal permission. These types of permissions are known as normal because of the resources they access, pose very little risk to your personal privacy. To put this in perspective, one such example would be an application connecting to the internet. No personal data is accessed and many apps need an internet connection to function properly.

The second type is what’s classified as a dangerous permission. These permissions are classified as dangerous because they access sensitive resources. These permissions could include accessing your friends contacts or from hardware pieces like your phone’s camera or microphone.

Our research motivations were centered around several key questions, however I think the bigger questions we want to address is how a certain permission use just changed? Is the app community getting better at protecting user privacy?

So just circling back to our main question, what factors may influence permissions in practice? We examined four variables to find out. We included application characteristics such as popularity, category, and ranking of the app; elements of app source code which included normal and dangerous permissions; culture due to varying attitudes on privacy; and finally, privacy reforms to protect the consumer’s rights.

But to put our topic in perspective, why do we care about what may influence permissions? In the app environment, many studies have indicated that several active applications operate as overprivileged. And what I mean by overprivileged is, in many applications access phone resources that are not within the context or purpose of their use.

There's also several types of advertisement libraries and bid in applications that track consumers behaviors. These libraries inherent a set of their own dangerous permissions and also act as overprivileged. Even worse, no permission is provided to the user when these permissions are granted.

We can magnify these problems with a case study. Take for example MoChat. MoChat is a clone application in which you run social media accounts simultaneously. In our study, MoChat was found to be our outlier. We found over 400 permissions in the applications source code and many of these permissions were found to be dangerous. You can see several of these permissions on the right side of our picture. And while some of these permissions may be contextual with any applications purpose, 436 permissions is enough to warrant a reasonable amount of concern to the consumer.

We then took a look at MoChat's previous privacy policy before the enforcement of a GDPR. In one section it stated that it does not collect certain PI such as user location data, however in another section it communicates that it does in fact collect session data with location. It follows up by stating that they are not responsible for harms and systems of hackers or virus attacks. So I think if we connect the dots here, we have an application that's collecting data that they say they aren't and in any instances of harm they are not responsible. Perfect example of how notices companies provide can be deceptive in practice.

And so the way in which we found applications like MoChat involved the collection and parsing of two data sets at two different points in time. We first collected over 4,000 Android applications in three separate countries. These countries were the United States, South Korea, and Germany. We chose these countries because their cultural attitudes and legal differences centered around privacy. We then followed up with a collection of the same 4,000 applications post GDPR to compare permission request occurrences.

We've analyzed privacy reforms among our targeted countries to see if changes in the law affected permission request rates. We noticed higher concentrations of privacy laws passed and drafted between the years 2015 to 2018. We also found that Google introduced newer guidelines and time periods mentioned. We wanted to take a look further and assess if the app community had been directly affected.

And so after analyzing our data set with policy, we found mixed results. Some were surprising and others were not so surprising. Take, for example, our results for collective permissions requested. If we take a look at the graph to our left, we see a gradual increase among all permissions requested at a rate of 9% GDPR post enforcement. We also see the gradual increase among both dangerous and normal permissions requested a rate of 3% post GDPR.

This gradual increase is also illustrated in our permission trend graph to our right. The graph included three of the most popular applications found in lifestyle, social, and age five and under categories. We can likely attribute these gradual increases with advancements in application technology. Take, for example, Facebook in our right graph. We see between the years 2014 to 2018 newer features being introduced to the Facebook platform, which would have resulted in more permissions requested in this application.

We next compared the top 540 applications for the Google Play Store and three different categories. We have reviewed applications in lifestyle, social, and age five and under, and found that permission requests were also increasing at a median rate of 6% post GDPR enforcement. We also found that both normal and dangerous permissions were also increasing in all three categories, with social and lifestyle applications increasing at the fastest rates. These results were also less surprising as many of these applications add newer features over time.

So the main takeaway here is that with each technological advancements in applications, we are allowing applications to access more personally sensitive resources at very gradual and at noticeable rates. And the implications of what we are sacrificing are far less apparent.

In our next results we found something much more interesting. In a condensed data set of the top 200 apps and age five and under category, we found that dangerous permissions decreased significantly since the enforcement of the GDPR. Overall, there was a 17% decrease in dangerous permissions requested, and among our countries, a never surprising result was exhibited.

Since the enforcement of the GDPR, South Korea's dangerous permissions decreased at a staggering rate of 26%. We also found that all other countries were also decreasing their dangerous permissions requested significantly. Whether these results are linked to privacy reforms such as the GDPR is difficult to assess. However, the countries we overviewed are in the process of reforming their data protection laws.

Take, for example, South Korea. Since July of 2018, South Korea has been in the process of modelling their data protection laws to be adequate with the EUs data portability standards. Similarly, states like California reformed a regional privacy laws to stem similar standards of the GDPR. And as many of you know, the GDPRs penalties can be significant. It's possible that regulatory enforcement is making a difference with data minimization policies, but then again it's difficult to demonstrate.

And while dangerous permissions have decreased and apps aimed at vulnerable populations, as a whole they are continuing to gradually increase. We can continue to see this increase observed further with our comparison model to our right. The frequency rate at which dangerous permissions are increasing at minimal rates with some permissions, while others are increasing at more significant rates. Among those permissions, the ability for an app to read your external storage on your mobile device has increased by 15% since the enforcement of the GDPR. Additionally, the ability for apps to access your phone's microphone and camera are also increasing at a median rate of 10%.

So just to quickly summarize what we have observed, collectively, both normal and dangerous permissions are gradually increasing over time. This is likely due to application advances and market demands.

[Whoops. How do I go back? Never mind. Yeah, sure]

Well anyway, dangerous permissions are also in the most popular categories for children age five and under are also decreasing at significant rates, with some countries seeing decreases at more significant rates.

And so the greater implication here is that with each year we provide applications of more access to sensitive resources on our phones. And as more devices are integrated of IoT technology, this fact could have unintended consequences for consumers privacy.

And so finally the bigger question, has recent privacy reforms made a difference? At this time we have limited evidence to state definitively. More analysis is needed and we are in the process of completing more data compilation to this work. A more updated version of this research is set to be presented at the Research Conference on Telecommunications and Internet this coming September. If you'd like to learn more, you can also reach out to me directly or on our research page at spice.indiana.edu. Thank you for your time.

[APPLAUSE]

MARC EICHORN: All right. Thank you. We'll now turn to questions. We invite questions from the audience here, or if you're watching on the webcast you can tweet us your questions to @FTC using the #privacycon19.

So thank you all for your presentations. I'll just get started with a question really for all of you, which is that we talked a lot about transparency on this panel and as you all know there's a lot of critics of notice and choice. And I think Kassem talked about the study that it would take you years of your life to read all these privacy policies and so forth.

And so I guess I just want you to expound on sort of how can we use these tools or this research to maybe make it easier for the consumer to sort of parse out information, maybe without spending so much time or maybe use sort of third party tools or automated tools to help do the job for them? I mean, Justin's work was largely based on review of these policies, so obviously transparency makes a difference and it makes a difference to us as regulators. But, I'd be interested in your thoughts as to the value to consumers.

KASSEM FAWAZ: Yeah. I mean it's a measured question of how we design the interfaces. I mean policies as natural language text documents have not been very successful as it said in presenting users with all of this privacy practices. So what we've been trying to look at is— and there's a great research, especially from CMU on how to do privacy nudges.

So what we've been looking into is, OK, now we have automated privacy policy analysis. Can we summarize this privacy policy into privacy nudge or can we start designing more intuitive interfaces so that when users are interacting with the website for the first time, when they're downloading an app, when they're shopping for a smart device, can we just show like some privacy nudges that would make users behave in a more privacy aware manner?

I mean a big problem in all of these tools is whether users are going to use them. And another problem is that where do we install these tools in the users decision making process? Let's say you're downloading an app or visiting website for the very first time. Even if you have a very nice interface that you have developed, it's unlikely that you're going to interact with it because your main task is visiting that website or opening that service and you don't want to be interrupted and your attention span is too little and too limited. So it’s really a question, how do we design these interfaces? That’s the first thing.

The second thing is that what information should we show to the users? And that goes back to the expected versus unexpected behavior. If I’m using Google Maps, I expect Google Maps to be collecting my location, right?. So I don’t really need to tell the user that, but I would need to tell the user, like if say some company is like accessing this information and selling it to someone else. So that would be an unexpected behavior.

There’s this very famous Android app, like a flashlight app, that’s accessing your location. That’s an unexpected behavior. So I mean we have to be strategic in what information we need to show to the user that might be of an interest. And how do we show that information through more intuitive interfaces that do not interfere with the user’s desks? So I think this is where we should be going.

MARC EICHORN: Yan.

YAN SHVARTZSHNAIDER: Yeah, I largely agree, and just kind of put our work in a bit more context. No pun intended. It’s all contextual really and I think we should change around how we look at privacy policy. At the moment privacy policies are put together as companies are trying to— you know on one side say we are privacy aware and we care about that, but then we’re really kind of user centric. It’s more about for the legal purposes and make sure that they are all covered and all that.

But if you change it around and you feel what Kassem was saying, privacy expectations, and those are changing based on the context. In our work, what we’re trying to do is to essentially find ways where we can tell the user that currently your privacy expectations are not aligned with the practices of particular service. So if you’re using Google Maps, you have some privacy expectation in mind. If the flows generated by that service basically deviate from those expectations, you should know about that.

And building on that, you can essentially provide a way where the user that interacts with the service is confident enough in order to make sure that the practices essentially align with what they expect. And if you build those, that kind of philosophy into the work, into the privacy policies— and be explicit because really what I was saying is the five parameters, it’s so simple, but it’s really not trivial. But it’s like sender, information, recipient, subject, and what are you going to do about it?

And if you are explicit about it, you can build on that. You can build tools to essentially either automate the way you check, which we already done in terms of whether it deviates from your privacy expectations, or visualize it in a more natural way for the user to see if there are anything that are not expected.

So I think that we should change. It’s a bit of a patchwork at the moment. We’re trying to introduce new attacks, tried to have flashing pop-ups. But really the attitude has to be around user and consumer, whoever is using the service, to see whether their privacy expectations are aligned with the practices of this service.

MARC EICHORN: Anyone else?

JUSTIN BROOKMAN: So I would just say, there are a few problems. One, there's no clear transparency obligations, at least in the United States. You had to have a privacy policy that does not require to say anything interesting. And so there's just no obligation of Turk folks to say what's going on? GDPR may be designed to try to address that, but GDPR is not actually being enforced. And as we saw from some of the presentations today, it hasn't changed behavior as was intended, or even as the text of the law would do.

The other point I was trying to make when I was rushing toward the end is about for whom are private policies to be made? I think the idea that they're going to meaningfully inform consumers directly I think is crazy. But I think the audience for privacy policies should be folks like the privacy account researchers, the FTC, the press. And so I think having an obligation to actually make them more detailed, to make them longer for a certain threshold a company, that they have like filing applications that make it more like an SEC type filing for sophisticated investors—look— I mean maybe not someone buying an index fund. —then I think the information could be digested and could be useful. And some of the problems that Yan identify as like parameter bloat, it's not for one person to try to keep in their head all the time.

So that I think might help somewhat with the transparency and external accountability issue. But more fundamentally, I think law needs to accord data sharing and collection behaviors with reasonable expectations and not to put too much onus on users. I think that's where the bulk of the law should be there on constraining behaviors to the context of the interaction. Transparency should be a back end to kind of make sure things aren't going wrong so that people can be held accountable by the Federal Trade Commission or others.

MARC EICHORN: Anybody else? So let me I guess just ask a follow up. So one thing, Christine, your research showed that privacy policies are getting a lot longer. And Yan, you looked at Facebook's privacy policy and found that it too was basically longer. So again, sort of picking up on Justin's point, that might be good from the standpoint that there's arguably more information for regulators or researchers and so forth. But overall, again, from the consumer point of view, if you are thinking of consumers as the audience, it's not necessarily a great thing. But do you have comments on sort of whether longer policies are more comprehensible to people or sort of just more complex and not understandable?

YAN SHVARTZSHNAIDER: All right. OK, I'll start. So as I was saying, more information doesn't actually lead to— you can put a lot of information there and will be in places that a consumer won't read or they won't expect. So, the placing is important, how you structured them and how you structured the statements themselves.

So what we were looking for our research is to see whether the privacy statement include all the relevant CI parameters. Who is this sender of information? What's the information about? Who is the subject, and so forth? So, actually the answer to your question is no, actually adding more information doesn't lead to— you have to be strategic about it anyway. If you are open about your practices and you would like to see the consumer understanding them, you will outline them explicitly and then, yes, you're right, they're going to be a longer privacy policy because now you're more explicit about it.

That's what companies will say. They say well, now we're making it even longer. So what do you want? Consumer won't read that. Well, then we'll build tools to make it easier, to either visualize only those that are not aligned with your privacy expectations— and in our work we essentially collected those; we crowdsourced consumer's privacy expectation, particular context, like with smart homes. And we can see if something is deviating, OK, your flashlight is sending your location information. Maybe that's not the app or the service you should use. And that's kind of the more meaningful and informed way of doing things.

And the other side just to conclude, is that yes, so maybe the consumer doesn't have the spam, but you guys have, you have the ability to go through that. And if it's explicit we would provide you with the tools to make a way to filter out the ones that actually deviate from maybe some additional policy or regulation, and then highlight those that those information flows that essentially are not aligned with what do you expect?

MARC EICHORN: Christine.

CHRISTINE UTZ: Yeah. So while we were doing this manual analysis of the privacy policies because the tool hadn't found one of them on the website, we sometimes came across policies that were spread out across multiple short sub pages. And usually they had some introductory side. And those types of policies usually try to explain the data collection practices in an easy to understand language instead of the typical legalese. And we didn't look into if some restructuring that had taken place during 2018 or during the course of our scans— but if I could make a wild guess I would just say that I assumed this had looked different at some point prior to GDPR. But this would certainly be an interesting thing to look into if there has been some restructuring going on pre and post GDPR.

JONATHAN SCHUBAUER: Just a comment. Show of hands in the audience, how many of you have actually ever read a full privacy policy? Anyone? OK. That's a surprising result. Well, I was going to mention, I think one thing about privacy policy is just getting longer. I think what it does do is help internally companies evaluate their collection policies. But what I would say is that for most users, you only have a couple seconds of their time.

And so I do think that does improve consent and notice and transparency in some regards, but I think the bigger problem is just getting the user's attention at a quicker rate. You only have about three to five seconds if anyone's really going to read those at all. So I think a better way of maybe fixing the problem of consent and transparency is maybe in applications, providing those consent notices in a smaller bulletin point form and in a way that they can read it and it's more accessible.

JUSTIN BROOKMAN: One quick follow up point. setting aside privacy policies, it's more to note that short form notices are terrible too. I mean like the GDPR consent experience, surfing the web in Europe, is— we did not land in an optimal result. And Christine's excellent chart showed that the vast majority of those notices are flagrantly illegal and not complying with GDPR. But even if they're optimal, even if they were like designed by the best experts, it would still be a bad experience. Every single website you need to go through like a bespoke expression of your preferences.

And so ultimately you don't want to have to do that, rely on so much notice. However the law does it, whether it is through like a browse or setting like the Do Not Track, or as just the law says the data needs to stay in context. Again, pushing people to privacy policies is absurd but again, requiring user engagement on every single privacy option is also bad as well.

JAMIE HINE: So Justin, I want to push you a little bit more on that comment. So I'm sort of paraphrasing a couple questions from the audience. Please keep sending them. These are great. So one of the questions we have sort of asked about how we can create more people centered policy? In other words, how can we ensure that people's expectations of privacy are consistent with what they're being told? And then another question asks whether the idea of nutrition labels for privacy policies was dismissed too soon, and sort of comment that nutrition labels have evolved over time and don't seem to be deterring innovation in food processing. So why can't we do the same with privacy policies?

So I guess what I want to ask a little bit more about is how do we sort of find the balance? So we have longer privacy policies and we're using automated ways to sort of interpret and determine what's inside the privacy policy so that we have more information, more long form privacy policies. And then Justin, you're sort of suggesting that short forms aren't working either, and Christine, some of your work suggests that in Europe there's a lot of changes by country and by language, and a lot of countries in Europe are sort of lagging behind others. So have we sort of lost the consumer in all of this? What do we do to provide more information to the consumer in an easily digestible manner?

KASSEM FAWAZ: I think we lost the consumer the moment we started calling them users on tech platforms. So I think at the moment consumers were lost. A comment about the nutrition label approach, it wasn't dismissed because the UI was bad or anything. The reason it was dismissed, that it requires someone to standardize their privacy policies this way and the only way they are done that way in food industry because someone says that food has to be reported this way.

So there was a legal requirement that actually someone has to standardize how nutrients and how other carbohydrates or whatever have been proposed this way. So the reason it was dismissed, because it's required manual effort and required someone to actually do it, and nobody was willing to do it and there was other standardized approaches for privacy policies and other notice that were dismissed this way now.

About the question what's the best way to present privacy policies, I like Justin's idea that we still need these long detailed privacy policies for compliance or for accountability purposes, the same way that SEC findings or whatever. There should be a document where everything is explained. So someone just really wants to look, they are there. So those long documents shouldn't go anywhere, they should still be there. But those are not what we should be showing to the users.

Now, we come to the other questions. What do we show to the users? I would like to have an answer to that, but I don't think anybody still has an answer for that. I mean short notices don't work. I mean I think what we should be going at, as what I said before— and there's research on this. —like nudges. We should be nudging users on more privacy-aware behavior. I mean we can't expect users to express privacy preferences for every website they visit. I mean, it doesn't work.

CHRISTINE UTZ: I think the main problem here is much more fundamental because right now what makes those services collecting vast amount of user data so attractive to users is that they're free and they're just convenient to use, which, of course is because convenience often requires personalization and data collection. And so there really is no incentive for companies to not collect vast amounts of data.

So in another context, I’ve once heard that this wasn’t in the security context, that the new paradigm should be let’s make security the most, the secure option, the most convenient to use. And so with regard to privacy you would say let’s make the privacy protecting version the most convenient to use. But I don’t see how this is going to work because you have the underlying incentives and the current business models of web services.

And actually I think about this business model issue a lot, but so far I haven’t really come up with anything that could replace the current business models other than paying for those services. And I see that this will be hard to sell to people. And we also don’t want a two class system where one set of users gets the more private version, but pays for it, and others who can’t afford it or probably just don’t want to pay that, they just have to deal with the more privacy invasion version of the service. So we need really some fundamental discussion about business models and if we can expect those to change and how they could change.

JUSTIN BROOKMAN: I’ll say I think by and large, consumers don’t want a lot more information about privacy practices. When they go to New York Times, they’re not looking for a lot of information about the New York Times privacy practices. They want to know that they’re being held to account, and most important they want to know that they can trust it, that they go in there, things are cool, which is why I think the law needs to focus on conforming behaviors to reasonable expectations or conforming behaviors to what’s reasonably necessary to deliver the product that the consumer asked for.

On the permission labels, look, I mean obviously we’re doing something resembling information label. We’re distilling it down much more deeply to a score. And again, this is like not going to the New York Times, but like for your car or for a major purchase where it might be an attribute that you’re willing to price or you’re interested in some degree of digestible information.

I think it’s kind of hard. I mean I think with the nutrition label you kind of know vitamin C. Well, you think you know what vitamin C does, you know what sodium does, you know what fat is. Like you don’t necessarily know all of it, but you have a general sense. It’s just harder to translate the detailed practices.

Like, how do you convey Facebook’s Custom Audiences where people share hashed identifiers in a third party data broker cloud to provide matched ads to you? How do you put that into a permission label? Look, there have been efforts to do it. There are smarter people than me who are currently working at doing it. It’s worth trying to do and trying different ways to do it, I just think it’s really challenging.

MARC EICHORN: Let me ask–we got a question from Twitter focusing particularly on whether any of you in your research were sort of focused on or looked at health care information, in particular, and whether there were any special characteristics of privacy policies related to health information.

JUSTIN BROOKMAN: I’ll say not as part of this project, though a little bit. And so I think health apps have access to a lot of very sensitive stuff.

This is like a research project in process, even more so than the paper I presented on. So I don’t want to say too much about it. And there have been a lot of news stories recently around health apps sharing data, either with analytics providers, including sensitive data in violation of the analytics provider’s terms, potentially sharing aggregate information back with employers.

I think this is not a well regulated space; it’s one of the reasons that we’ve again push policy solutions. So in California and New York we’ve been pushing bills around wellness programs and wellness apps, limiting the data that can go back to your employers or potentially life insurers or other people who maybe you don’t expect to get it.

I know at the federal level, Senators Klobuchar and Murkowski introduced, again, a pretty thoughtful law to try to conform health apps, which fall outside of HIPAA, to have some more HIPAA like rules around it because I think that there are a lot of dubious practices and not a lot of clear norms that apps or wellness programs are clearly following.

MARC EICHORN: I guess let me ask another question from the audience. The question is, should states step in to require apps to limit data collection or coverage? So I think this goes to sort of this idea of, well, instead of just being transparent about practices that may be whatever, maybe you go directly to sort of limit certain practices or uses and then perhaps you have fewer things to inform consumers about, or maybe you again only inform consumers about things that they may not expect.

YAN SHVARTZSHNAIDER: I’ll start off. Just a quick kind of reminder, when I was talking about contextual integrity as a theory, basically said some very fundamental thing, privacy is not about secrecy. It’s not that you don’t want to share, it’s not that you want to just keep it all in, and you want to use those services and just we want to make sure that whatever information you share is shared appropriately. There is an appropriate way of appropriate information flow.

So I’m not a lawyer and I think this idea might be interesting that states can take more charge of that. But just on pretty fundamentally, I am not a fan of the idea that we just lock everything in and then see what happens. It’s really working hard and finding out ways and using some of the things we are designing and thinking about, is how to ensure that there’s an appropriate information flow because we do like those services. We think they’re contributing. They’ll help us do a lot of things. And so as long as they do that in appropriate manner, then it’s fine, it’s just you don’t want to stop them from operating. You just want them to do it in such a way that aligns with the words and privacy expectations. It’s not an easy thing, but it’s a challenge and I think we are on the right path to actually solve it.

So I’ll be more optimistic than other panelists. I think we’re not there yet for sure. But to the type of work I heard today and the things that we are thinking about, I think it’s doable, we will just have to again maybe change our perspective, focus on the consumer more and see how we can help. And I think that we can do it. Yeah, I don’t think we should go drastically, just stop collecting. Yeah.

JONATHAN SCHUBAUER: And just to weigh in, since it’s address in the app space and billing and state and making the rules and data minimization policies and things of that sort—with applications that access dangerous resources such as what I’d mentioned in our talk with dangerous permissions, some of those are contextual and need to be used in applications.

You can take, for example, a lot of you probably use Snapchat and it accesses your record audio and your camera. well, an Android, that’s considered a dangerous permission, but that's also contextual. So if we lay out these rules that limit our developers practices, that creates problems. So it's a challenge to get the law in the right spaces with, I guess, with the data collection, but there, again, we're not quite there yet. So I guess being aware that we do need certain practices and we can't limit all of them is something to think about.

JAMIE HINE: Jonathan, if I can follow up. Your research focused a bit on this concept of dangerous permissions. And I was wondering if you had looked at all over how dangerous permissions have changed over time, and I'm curious whether there's a role for the app stores to better sort of fine tune the permissions. In other words, are the permissions just collecting too much information? Should they be finer grained to be collecting more limited app information to sort of accommodate this increase in features over time? I don't know if your research will do that.

JONATHAN SCHUBAUER: Well, it's not centered entirely around that, it's just really looking over the occurrences and how they've changed over time. But one way I guess that Apple has tried to address this issue is using a model of how they ask these permissions.

They're using Ask on First Instance I believe, or I might have gotten that acronym incorrect. But basically the way it works is that it would request a dangerous permission notice when it's in context. So I'll use the same example with Snapchat. When you have a dangerous permission such as your camera— applications used to introduce dangerous permissions right in a row. If you downloaded an application you'd get a list of messages and you'd hit Allow, usually within the first three seconds. But with something like Snapchat when you're using a camera, it'll present a notice and you'll hit Allow or Deny. It needs to use the camera, so it's in its context. So just certain aspects like that can definitely improve that environment.

KASSEM FAWAZ: Yeah. I have a follow up on this. Android had this laundry list of permissions a few years ago. They have changed their permission model to be runtime permissions now. Experience with Androids, they have a permission for each sensor and for each information access, which works somehow well.

But what Apple has done is they have added another dimension to it, like foreground versus background information access. And I think that has been great in terms of granularity. You expect Snapchat to be using its camera when it's on the foreground, but you wouldn't like it to be using the cameras in the background. So this another dimension is actually pretty useful in terms of privacy protection.

JUSTIN BROOKMAN: In general, I'd like to see the app platforms do a better job of policing because I think in some of the research we saw there's tons of apps who, hard to say for sure, but like fairly strong indications that they're in violation of Android's terms, and I know the FTC has gotten some complaints around a lot of companies clearly in violation of COPPA on the app platforms. And I think the app platforms could do more about it.

Kassem's right that I think the platforms have constrained some of that kind of over time. And they've kind of seen it to be user platforms had done some work to at least improve the policies, if not always the enforcement, but also limiting some of the technical things companies can do. But I mean in general, I think app stores beyond privacy— I mean look at a lot of the kids apps. I think there's this kind of a broader conversation about the role of platforms and content moderation, but I think that companies need more legal incentives to take responsibility for what happens on their platforms.

JAMIE HINE: Well, with that, our panel has come to an end. I want to thank the panel, and please give our presenters a round of applause.

[APPLAUSE]

And with that we’re going to take a short break. Please be back in the auditorium shortly before 11:20 when the next session starts. But I want to let folks that are in the room know, the cafeteria is open for just about 10 more minutes. So if you’d like a coffee or a refreshment, head over there quickly. So we’ll be back at 11:20. Thank you.

[MUSIC PLAYING]

ANDREA ARIAS: All right, everyone. If you could please take your seats, we're about to begin our next session. My name is Andy Arias. I'm an attorney in the Division of Privacy and Identity Protection here at the FTC. My co-moderator is Yan Lau, an economist with the FTC’s Bureau of Economics. Our second session for today is Consumer Preferences, Expectations, and Behaviors.

You'll hear from five researchers. Their presentations will be approximately 10 minutes or so. We'll conclude with about a 20 minute, maybe even longer, discussion where we'll identify common themes and ask the presenters about their work and its implications. We won't be asking questions until after all five researchers have presented. However, feel free to start writing your questions as the presentations are going on. Raise your hand, we'll hand you a comment card, and they'll bring them up to us. Or if you are on the webcast, please go ahead and tweet us @FTC #privacycon19.

Without further ado, I'll briefly introduce our presenters. Their very impressive and more fulsome bios are on our website, or if you're here in the room we have some copies of their bios up front as well. First on my left is Katie McInnis of Consumer Reports; to Katie's left is Mahmood Sharif from Carnegie Mellon University; to Mahmood's left is Noah Apthorpe of Princeton University; to Noah's left is Kristen Walker of California State University Northridge; and finally, though certainly not least, we have Yaxing Yao of Syracuse University. Katie McInnis will start us off with her presentation on her historical review of consumers' privacy expectations and how that aligns with Consumer Reports’ latest national survey.

KATIE MCINNIS: Hello, and thank you to the Federal Trade Commission for the opportunity to speak with you today. At Consumer Reports we're developing a historical report on consumers awareness and responses to online tracking techniques from around 1995 until today. I'm excited to share with you today some of our initial findings.

In an order to put together this report, I review publicly available nationwide surveys about consumers understanding responses to online tracking techniques, studies about what consumers understand about these tracking techniques, and historical survey data that were conducted by Consumer Reports. My initial research indicates that consumers initially became aware of online tracking techniques through being able to identify some common tracking tools such as a cookie.

Over time, consumers have become more aware of online tracking techniques like pixels, cross device tracking, and the correlation of offline and online data to create a more full profile of the user. However, the increase in consumer understanding of tracking tech and their ability to mask their activities to prevent these trackers from knowing about their online activities and offline movements have not and could not keep pace with the evolving tracking technologies that have been developed.

We have reached a point where tracking technology has become too sophisticated for the average user to understand and therefore control. Who in here--and feel free to raise your hands--has had a friend, relative, acquaintance or co-worker, sincerely ask you whether or not a social media company is hacking into their phone in order to spy on the conversations to deliver them more personalized ads? And I'm sure that some of you have wondered that yourselves. You see that ad about something that you think that you only talked about in the vicinity of one of your devices, but never actually Googled or looked into.

As you can see from the number of hands in the room and the number of repeated online news reports about this, this discussion is live and well and it reveals that consumers don't understand how they're being tracked nor do they understand how personalized ad delivery works. This debate over whether or not companies are listening to you through your phones is probably one of the best, if not the best, examples of the real gap between consumer understanding of tracking technology and the actual tracking that is occurring.

Consumers are seeing a highly accurate and some cases creepily accurate ads. And without a good understanding of how their activities are being tracked, consumers are driven to this assumption that they must be listened to by their phones in order for these ads to be delivered, in order for some ad to know that they are afflicted with some sort of medical ailment, that they want to go on a trip to Greece, or they're in the market for a new fridge.

So how did we get here? Although most consumers didn't really use internet in significant numbers until the mid-1990s, the population of people who are currently debating how these ads can be delivered with such accuracy include not only what we call digital natives, but also people have been using the internet for over two decades. But just five years into regular consumer use of the internet, around 2000, consumers already had really strong opinions about the privacy of their data and the ability of these trackers to track them.

For instance, 86% of users in a nationwide poll that year said that they disliked this online tracking, and wanted websites to get opt-in consent before collecting this information. And 91% of users disliked the abilities of just regular trackers to know that they're going from site to site to site.

Despite these strong feelings about tracking, more than half of the users in this period, internet users, could not identify a common tracking technology which is a cookie. So 56% of people weren't aware of how they were being tracked, but they were feeling very strongly about it. So this divide that we see today, the divide that has continued to be involved in our discourse around online tracking, is really as old as consumers significant use of the internet itself. And although this awareness of online tracking and the techniques by which this is done has evolved over the last few yearsthe real change we see in the early aughts and late '90s is among consumers, it's not a push back on this tracking technique, but rather more confidence and assurance that the activities that they're using the internet for, such as looking for medical issues or connecting with friends or shopping, can be done well and with some kind of security.

By 2005 we not only see greater use of the many features of the internet, but we also see that consumers are feeling that they have some sort of rights over this data. For instance, in 2007, although 80% of individuals knew that marketers have some ability to track them across the web and 62% knew that a company can tell if they open a marketing email from them, 75% of those surveyed thought that the mere presence of a privacy policy on a website or an online service meant that a company could not share the information that they collected about them with others.

So there's this clear misunderstanding of what is happening here. People are being tracked, and yet they feel confident using the internet and they don't primarily understand the documents are meant to tell them about their privacy rights online. Despite this discouraging statistic, the aughts also feature one of the first instances of what we now call a tech lash, with the revolt against a Facebook beacon program.

The Facebook beacon program (for those who are unfamiliar) sent data from external sites to Facebook for the purpose of targeting ads and allowing users to share their online activities with their connections on the social media site via updates in the news feed. And although the change to the privacy policy on Facebook happened earlier in 2007, it received almost no notice.

From the beginning of this Facebook beacon program, it was mired in controversy. Protest groups formed quickly on the Facebook site itself and at least one major retailer, Coca Cola, dropped their use of the beacon program. Although we know that Facebook eventually changed its program to an opt in system and shut it down within two years, what Facebook and thus other online retailers learned is that you must obscure the tracking that you're doing from the users that are being tracked. Obscuring those tracking methods is essential to preventing consumers from knowing that it's happening, and therefore taking some sort of control even with the paucity of tools that they are left with.

And the Facebook beacon incident marked a turning point in our experience of online tracking technologies. As Bernhard Debatin and his co-authors noted, the beacon scandal was an accident because it made the users aware of Facebook's vast data gathering and behavior surveillance system. Facebook's owners quickly learned their lesson.

The visible part of Facebook, the innocent looking user profiles, and social interactions must be neatly separated from the invisible part. As in the case of an iceberg, the visible part only marks part of the whole. Consumer awareness of being tracked grows during the period of 2010 and 2015. During this time we have a couple of news moments that make consumers aware of what this data can tell and reveal about them.

For instance, we have the now infamous example of Target revealing a young woman's pregnancy before her family knew due to data that they had collected about her. And we also have the Snowden revelations which led consumers in the US to be aware that both commercial and government actors were collecting vast amounts of data about them. And by 2016 to 2019 during this period, consumers are very aware that their activities are being tracked. And a Consumer Reports survey in this period noted that the majority of adults not only are aware of this tracking, but also aware that digital profiling may occur without their knowledge.

Users during this period are more skeptical of tech companies and concerned about the information they share online. In addition, more individuals report using privacy protective tools like ad blockers. Currently, users know about the existence of tracking, have a little bit of awareness of what the methods are to track them across the web, but they really don't fully understand the myriad ways in which they are tracked. Furthermore, they're left with few controls to control such tracking.

For instance, a survey in 2018 found that a quarter of Americans use the Do Not Track signals to protect their privacy, despite the fact that most websites and other online services do not observe this signal. And in addition, many of the techniques consumer education outlets like Consumer Reports have been recommending to their users to protect their privacy since the late 90s and early aughts remain the same.

Examples of some advice are giving false information, changing your default settings, blocking third party cookies and clearing your cookies often. Although we've added recommendations to this list like using a virtual private network, consumers are really being told to do more and more to protect their privacy. And these advice steps are taking more time and effort, and in some cases, taking also money. Despite this growth in a more time intensive list for best practices for consumers, even if you use all of these tools some amount of data will still be tracked and used to create a profile of you for marketing purposes.

So although over time we do see greater awareness of tracking, that comes as a result of consumers increased experience across the web, consciousness-raising moments that either come through personal experience of tracking or news reports about tracking, and consumer education efforts. Awareness that you're being tracked alone does not lead to greater consumer control over and protection of their private information. We are left with an environment where consumers know they are being tracked and are largely unaware of how this tracking is done and unable to control such data collection.

So what do we do to bridge this gap. We need to take the onus off of the consumer and we need to put rules in place that prevent over-collection of data. We would need to allow for consumers to have access, deletion and correction rights to their data, and they also need to have stronger security measures and easier to use and understand tools. Although tracking what consumers know and understand about tracking, and ad tracking is helpful in understanding how our mission creep of data collection has progressed, transparency around these tracking methods alone is not enough. We need transparency in addition to better tools and stronger rules and regulations around which tracking can be done for consumers to protect their highly private data. Thank you.

[APPLAUSE]

YAN LAU: Thank you, Katie. So next you'll hear from Mahmood about valuing privacy in hypothetical versus realistic scenarios.

MAHMOOD SHARIF: Hello, everyone. So today I present our work on comparing hypothetical and realistic privacy valuations. This is ongoing work with my collaborators from CMU and UMD. So when we talk about privacy preferences, we usually refer to people's willingness or comfort at sharing personal information, and studying privacy preferences is important for different groups.

So for example, system designers often build systems that collect data about users. For them, it's important to know what kind of data is OK or not OK to collect. Also, users often share their data with systems in return for certain values. And in such cases, it's important to know what's the minimum value that users expect in return for their data.

Studying privacy preferences is also important for policymakers. For example, user's data is often compromised in data breaches. When this happen, it's important to be able to measure the loss. Also, it's important to know if there are certain kind of data sharing that consistently violates the privacy preferences of users. For such data, policy can take an active role to prevent or to incentivize sharing or collection by services.

So measuring privacy preferences is important, but unfortunately it's also challenging. And one reason that it's challenging is that privacy preferences are very contextual. And so when we measure privacy preferences, it's important to take the context into account because privacy preferences may be different under different contexts.

So for example, users’ willingness to share their personal identifiable information may depend on how it's going to be used. Another way that we might go about measuring privacy preferences is by asking people how much they value their personal information. However, such valuations may be subject to certain biases as well. So for example, in hypothetical scenarios, people might be affected by hypothetical bias, which may lead them to overestimate certain values.

Another reason that studying privacy is challenging is the privacy paradox. And here, what prior work has shown is that people often don't act in ways that align with their privacy preferences or with their stated privacy preferences. Or, in other words, often people don't take an action to protect their privacy and end up sharing their data freely even if they previously said that their data is important to them.

And so in this work, we want to see if we can predict privacy valuations. And by privacy valuations, we refer to willingness to sell and selling price for personal information. And more concretely, we want to see if price valuations can be determined by three different factors: the attribute type, the receiving party, and the scenario realism. And while prior work looked at how each of these factors individually or in pairs affect privacy preferences, we want to see how all of these factors combine to affect privacy valuations.

As a second research question, we want to see if hypothetical bias can help explain the privacy paradox. Or, in other words, we want to see if the privacy paradox can be explained by those cases where the privacy attitudes or reflection of preferences in hypothetical settings. So to answer these questions, we ran our online study and we recruited 434 participants from Prolific, a crowdsourcing service that's similar to Amazon's Mechanical Turk. And participants in our study were asked to provide the minimum dollar amount for which they would be willing to sell their personal attributes, or they could decide not to sell at all. And the context for selling was an information marketplace that's operated by CMU. So you can think of it as eBay, but for personal attributes, and participants would be selling on this marketplace.

We wanted the valuations of the participants to reflect the actual worth that they place on their personal attributes. And so we collected the selling prices through an incentive compatible auction mechanism, and this effectively prevented the participants from gaming the system--for example, by assigning a higher value than the perceived worth just to make extra money.

So here's the format that we collected the selling prices under. So participants sold this where we would ask them for how much would you agree to sell your attribute to each one of the following parties? And they could decide to sell or not to sell, and if they decide to sell then they had to provide a dollar amount.

We asked about seven different attributes, including less sensitive ones like age, and more sensitive ones like phone number. And participants could also provide different values, different selling prices for different parties, and we asked about six different parties and tried to capture ones that would use the data in a variety of ways, as prior work has shown that this affects privacy preferences.

In the main experimental condition in our experimental factor in our study was the realism of the marketplace. So participants would be assigned to different conditions that would affect the description of the marketplace that they saw. In one condition, the marketplace was described as real and functional. In this condition, we asked participants to sign in with Google single sign on in order to collect different attributes about them, things like home address and age and phone number that were attached to their Google profiles.

In the other three conditions the marketplace was described as hypothetical. And the level of hypotheticalness varied from an operational marketplace, the concept that we were interested in testing, to a completely imaginary marketplace that's used only for research purposes.

So now I'm going to present the results. And the first result is that attributes that could be used to contact the participants were generally sold for higher amounts. So here, you can see the attributes on the x-axis and you can see the selling prices on the y-axis when selling to political parties. Different lines correspond to different conditions. And you can see that for email, home address and phone number, the prices were generally higher than the other attributes.

Another interesting observation is that the selling price is dependent on who was buying. So while the prices tended to be lower for research pool and federal agencies, they were generally higher for the other parties. And it's possible to infer that, at least for some of the participants, they wanted more money in return for their data if it was going to be monetized or used to serve them ads. This finding aligns with prior work.

So given that other researchers have observed the privacy paradox, we expected that the hypothetical valuations would be generally higher than the realistic valuations. Surprisingly this wasn't the case. In general, the hypothetical valuations were comparable to the realistic valuations, however there were two exceptions for phone number and home address, which were sold for $5 to $3 less compared to others.

We also looked at what calibration factors may apply to the different attributes. And here calibration factors are defined as the ratio between the average hypothetical price and the average real price. And because of hypothetical bias, we usually expect calibration factors to be larger than 1. In our case, we found that most of the calibration factors were close to 1 and that the largest calibration factor when selling phone number to ad networks was around 1.6.

These calibration factors are actually quite low. So for example, List and Gallet tried meta-analysis in economics, and they found that for public goods, things like access to recreational parks, the calibration factor was around 4, and that for private goods, things that you might be able to buy at the store, the calibration factors were around 8. Lastly, for the likelihood of selling, we did not see any statistically significant difference based on the scenario realism.

So going back to our primary research question, can we actually predict valuations based on scenario realism, attribute type, and receiving party? Here we'll look both at dollar values and the attributes rankings, or in other words, the relative valuations. For dollar values, turns out that the answer is not yet. And this is mostly because different participants had vastly different baseline selling prices, however if you know someone's baseline then it's possible to make accurate predictions.

For attribute rankings, turns out that the answer is yes. And this is mostly because attribute rankings were more stable than absolute values. So in fact we had the same average ranking across the conditions regardless of realism or who the receiving party was. Also turns out that attribute rankings or the prediction of attribute rankings can be improved by eliciting a couple of rankings for a couple of attributes from the participants and using this in a machine learning model to make predictions.

So to wrap up, despite what we may think because of the privacy paradox, it turns out that it may be possible to predict privacy preferences in certain cases from hypothetical scenarios. And in contrast to other kinds of goods, privacy valuations aren't that affected by the hypothetical bias. So for example, in our study, we found that attribute rankings were stable regardless of the realism of the condition or the receiving party. So for a system that only requires an understanding of attribute priorities, those can be learned by asking a couple of questions in a hypothetical scenario. And while we still cannot predict the selling prices accurately, if you know someone's baseline then it's possible to do. And it turns out that the baseline is not a function of any of the factors that we measured. Thank you.

[APPLAUSE]

ANDREA ARIAS: Thank you, Mahmood. We'll now hear from Noah Apthorpe. He'll be presenting on his study measuring whether COPPA's regulations align with parents’ privacy norms.

NOAH APTHORPE: Great. Hi. Thank you everyone. So I'm here to present on research evaluating the contextual integrity of privacy regulations, specifically about parents privacy norms versus COPPA. And I emphasize that this is joint work with Sarah Varghese, who is a Princeton undergrad and is doing her senior thesis on this topic, and Professor Nick Feamster, all of us from the Princeton University Center for Infotech Policy.

So back in 2017, the FTC clarified its guidelines on COPPA to say that it included connected toys or other internet of things devices. So these include many products ranging from stuffed animals to tablets to robots that are all targeted at children under the age of 13. And what we want to know is whether the parents who are actually buying these toys for their children, whether or not the current regulation here stemming from COPPA mandates data handling practices that are in line with their privacy expectations. And then, of course, because we're academics we want to ask a more general question too, which is, how can we test broadly whether privacy regulations are aligning with the social and cultural privacy norms of affected populations on a particular issue?

So here we turned to the theory of contextual integrity. And as you heard about in the first session, this is a theory which restates privacy as the appropriateness of information flows in given contexts, and specifically the theory allows you to define an information flow using five parameters. And so what we did was we generated lists of these parameters all from the IoT toy context. And this allowed us to create permutations of parameters leading to over 1,000 descriptions of data collection practices. And these descriptions you can sort of think of as a very short story outlining a flow of information from a toy.

So here you can see some examples of what we came up with, with toys ranging from dolls to robots to walkie talkies, attributes like location, birthday, audio, and recipients including the manufacturer of the toy, and third party service providers. And then importantly, what we did is for the transmission principle we drew directly from the FTCs six step COPPA compliance plan, which specifies different requirements for online services that are targeted toward children. So it'd include things like if the owner has given verifiable consent, or if the product implements reasonable procedures to protect the information collected.

Once we had all of these descriptions generated, we got a panel of parents of children between the ages of three to 13 and we asked them to rate these flows on an acceptability scale. And here you can see one example question of the parents who were taking this survey would see. And in this case, it's about a toy walkie talkie recording when the child is actually using it. And then they can go through and they can see the various data collection practices and rate whether or not they think this is acceptable or unacceptable.

Once we've done this we end up with a fairly rich multi-dimensional data set and we can see how the different parameters, including the COPPAs specified practices, the toys themselves and the data themselves actually affect how people view the acceptability of these practices. So you can see here an example of the sort of data that we can find from this. And there's multiple figures like this in our paper. And I’ll walk through a few key results here, but I encourage you to go and look at the paper to see more in detail.

So first of all, the type of analysis that we do is to compare across parameters and also across participant demographics to see whether there are certain populations which may or may not be well aligned with their opinions and with COPPA. So first let's take a look at these two columns that I've highlighted with the blue outline. These are data flows which are either not allowed by COPPA or not given any sort of criteria that the regulation is placed on. And we see that on average those are viewed as being unacceptable.

Here if we compare against the remaining flows, which have been given a COPPA inspired criteria, those do tend to be acceptable. So the take away from here is that the types of practices that COPPA is requiring are generally in line with the sorts of things that parents are wanting. Things like security, consent, and the longer list of transition principles that we asked are having a positive impact on the views of the parents toward these flows.

However, since research in general should be viewed with criticism, I want to point out here in the gray bubble of skepticism that COPPA guidelines are fairly broad on these topics and things like if the privacy policy permits it. Well, we just had a whole session about privacy policies and the various challenges that they pose.

So even though we think that this indicates that the guidelines which are currently in place may be in the right direction, are sort of capturing the high level ideas that parents are concerned about, that doesn't mean that the actual implementations of the toys are actually meeting those expectations themselves. And even further, despite the good work by the FTC and the state attorneys general, there are a lot of smart toys which still do not meet COPPA compliance. And so just because those rules are in place, there's more enforcement work that needs to be done.

To continue, I want to look here at this set of flows which is all about notification and consent. So these flow descriptions included parameters about privacy policies, about verifiable consent, about being notified before data is collected. And we see that on average those types of flows tend to be more acceptable than alternative flows which include criteria of confidentiality and security. Things like if the data is being protected and encrypted appropriately when it's stored, when it's in transit.

And this I think was sort of surprising to us because we sort of felt that confidentiality and security may in fact be more important to protecting privacy than the notification. And aswe heard in this past session, a lot of problems with existing consent mechanisms. So here we have a sort of disconnect between what parents are reporting they believe to be important, which is sort of consent above all else. But there is this disconnect between that and what we see in practice, which is there are a lot of real deep problems with current consent mechanisms that need to be addressed.

And then finally a point of note which would be interesting to this crowd, we found that the percentage of participants who stated they are familiar to some extent with COPPA were more accepting of data collection in general. And this was statistically significant with the appropriate multiple hypothesis corrections. And we thought that this was potentially interesting because we would have assumed going into the project that participants maybe with a bit more knowledge about what's happening behind the scenes may be more concerned or perhaps no different. But it's the fact that they're more familiar, meant that they were more accepting, may indicate that there is some sense of security here which may not line up with the actual implantations of the toys that we're seeing in practice.

So what does this mean from a larger picture? Well, the takeaway that we have is that regulation can indeed help align data collection policies with privacy norms. I think this should be encouraging to everyone here because it means real meaningful steps can be done from the regulation perspective to push companies into doing data collection in a way that aligns with social and cultural norms.

Well, we also found that there are variations across demographics, across implementations, and across contextual factors like types of information, like how it's being collected and stored, and even the types of products that are doing it, which cause some fairly wide varieties in whether or not people view that as being acceptable. And these sort of smaller demographics, smaller considerations, are really important because when you're trying to regulate, it's often easier to paint a broader picture. And diving deeply into the variations that are coming up from these contextual factors we think will prove to be important and must be taken into account.

And so where do we want to go from here? Well, having shown that this technique can provide some insights in the IoT toy space, we want to take this to other regulation. I think that HIPAA's an obvious follow up to this study. And more broadly, we hope that other researchers, regulators, consumer advocates, will take this approach to see maybe in niche demographics or niche situations where there may be disconnects between what the public is looking at and thinking in their privacy reputations and what may actually be going on.

We'd also like to repeat this survey over time because the IoT toy space is sort of rapidly evolving and fairly new. We want to see how users norms change as internet connected products for children become more mainstream. And then finally we also want to follow this up with larger sample sizes from different sub-populations because it's important to ensure reliability and diversity of any sort of user-based survey study such as this one.

And so with that, I'd like to thank you all for listening, and especially like to thank some collaborators from NYU, Cornell Tech, and Princeton CITP who helped out with this project, and look forward to your questions. Thanks.

[APPLAUSE]

YAN LAU: Thank you, Noah. Next you'll hear from Kristen who will be talking about her study to determine what influences children's privacy behaviors, specifically when interacting with YouTube.

KRISTEN WALKER: All right. If I click, it'll come up? All right. Hello, I'm Kristen Walker. I'll be presenting work today with Craig Andrews and Jeremy Kees on the role of cognitive strategies, age and motivation in children's privacy protection, building on the lovely presentations so far.

So if it's not apparent to all of us already, there is incessant online activity. Most teens have access to a smartphone, almost half of them report that they're on all the time, and five to 15-year-olds report that they're online an average of 15 hours per week. Even preschoolers age three to four spend over eight hours per week online and parents report that they struggle with really how to manage this in addition to their own online use and time.

So we're really interested in how do we encourage children to protect their information online? Do they know what they need to do? So I'm going to take you through a little bit of the background that we don't have a lot of time for lit review, et cetera today. So the research questions that I'll discuss in a second came about because my colleague, Craig Andrews, who's here today, remembered that in the late 1990s sites such as Kids Com had internet safety quizzes on them. And in a sense, that was trying to encourage internet safety at the time. Obviously it's come a long way since then.

So we wanted to look at our children and teens and parents protecting themselves online. Overarching questions. Are there ways to empower children and teens regarding their online safety knowledge and behaviors? And then is it better to have children learn more safety themselves, privacy, safety, or have parents enforce privacy? What we base this on is a sharing surrendering information matrix. This is a conceptual matrix that is really designed to look at both trust, faith, and active and passive protection behavior.

What we focused on solely was the active or passive protection behavior. Do they place conditions on their exchange when they're exchanging information online? Can they be motivated to restrict sharing access, or motivated to be more active in their protection behavior?

So we're looking at precursors to persuasion, motivation, and ability. Here we're looking at motivation to restrict online information in general, and then enhancing ability using cognitive defense strategies to improve privacy knowledge. This is based on a lot of work from Brux et al. on cognitive defense strategies.

We used an educational video clip. This clip was an educational campaign that was funded by the Digital Trust Foundation, going back to the Beacon example earlier. And then we also used a quiz with feedback, and that content was aligned with the educational video content.

Overall our framework, it really describes the cognitive stages of development. And we used three cognitive stages of development for children: six to seven or the limited age group, the acute age group for 8 to 12, and as a parent of a 15-year-old, the one I find humorously titled, the 13 to 15-year-old strategic age group. We then looked at cognitive defense strategies. As I mentioned, a video and a quiz, and a control group, and then I'll take you through the rest of this a little in a couple of slides.

All right. So our predictions. The first one was really the quiz should work better for improving online safety beliefs, for improving the importance of restricting a YouTube video and improving willingness to restrict sharing access to a YouTube video. This will make sense when we see the methods.

The next hypothesis was children who are more motivated to restrict would have more favorable beliefs about online safety. They would think it's more important to restrict the YouTube video that they watched and show more willingness to restrict sharing access than those with lower motivation.

And then lastly, the strategic age group will have more favorable beliefs about online safety, think it's more important to restrict a YouTube video that's watched, and have greater willingness to restrict access than other groups.

So our methods. Data collection: we used actual children for this study, so the IRB process was very lengthy. We used an expert firm who had experience with children and teens and we went through a lot of IRB approval processes. This required double consent procedures, so the parents had to approve and consent, and then the children did.

For our main study after the pretest, we looked at 513 children and teens. They were randomly assigned to either read a quiz, watch the educational video, or neither of those. They then answered questions about their online information, opinions and beliefs, and then they watched a video online that the parents chose. And this was important for IRB approval as well.

It was a 3 by 3 between subjects design, looking at cognitive defense strategies, age difference categories, and high-low motivation as I pointed out. Our key dependent measures were online safety beliefs, the importance of restricting a YouTube video watched, the willingness to share, and if yes (they were willing to share), then with whom would they share that.

So briefly going over our findings with the time we have left, the first finding was that the quiz was significantly better than the video and control with online safety beliefs. The next was that with the importance of restricting the YouTube video watched, the quiz was significantly better than the control group. The age category effects were only significant for online safety beliefs. And here are the means; they're compared down. I won't take you through them due to time, but they're here for your reference.

So with willingness to share the YouTube video, and if willing to share if it was with everyone,we see that the control group was significantly more willing to share. We see here that the eight to 12-year-olds, the limited group, were significantly more willing to share the YouTube video that they watched. We see that the quiz was significantly better with audience restrictions as compared to the video. We're attributing this maybe to overconfidence, use of device experience. The strategic age group was significantly more willing to share the YouTube video as compared to the limited group. And again, we're wondering whether this is overconfidence or experience with the device use. And then a counterintuitive finding was that the more motivated they were to restrict, the more willing they were to share with everyone.

So discussion and policy implications, just to review. With online safety beliefs and knowledge, the quiz with the feedback, the strategic age group did well, both of those, as well as those with high motivation to restrict. Those turned out best for that. The decisions to restrict sharing the YouTube video, the video came out best with that, and strategic age groups.

The video we think was best because it actually demonstrated. There's a mice that represent companies and the mice gather up the crumbs after eating the cookies the kids leave. So we think that might have an effect. If they were willing to share the video, was it with everyone? Of those who were willing, the older ones were more, and the more motivated youth had the highest. And again, we're wondering whether this is overconfidence.

I think this is the most interesting for me as a parent. If the children perceived parental restrictions, then they showed more positive online safety beliefs and greater importance of restricting the YouTube video. So we think that parents do have an influence. I think that bodes well as a parent.

Policy implications. Really we're looking at COPPA perhaps making overt nudges for this kind of-- I think one presenter mentioned it earlier as a privacy nudge. We're really looking at whether or not maybe a national campaign can be created because obviously the educational quiz was best, but the video was also really useful. Maybe that can be combined with other websites, et cetera. And a good example of a national campaign would be the FDA’s Real Cost Campaign. Thank you.

[APPLAUSE]

ANDREA ARIAS: Thank you, Kristen. Our final presentation is from Yaxing Yao. Yaxing will be presenting on various models he's developed through machine learning to predict consumers' privacy behaviors in their smart home. Take it away.

YAXING YAO: Good morning. My name is Yaxing. I'm a fourth year PhD candidate in the School of Information Studies at Syracuse University. I'm very happy here today to present our most recent work on predicting individual users’ smart home privacy preferences. This is joint work with Nata Barbosa, Joon Park, and Yang Wang.

So first of all, the phrase smart home and privacy has been in the news very often. A study has shown that consumers are concerned about privacy and security in smart homes, for example, secondary use and proper appropriation of data collected through smart home devices. The fact that company have a commercial interest in user data can create an environment where secondary use and appropriation are commonplace inside a home, causing the home to become a place where privacy is no longer included, threatening the long settled notion that the home is a private and intimate place.

So these facts create an opportunity for the design of privacy enhancing tools that can help developers in respecting the privacy of smart home users and build user trust. Our goal of this research is to enable developers to derive actionable steps towards respecting the privacy of smart home users in a personalized way.

So how are we going to do this? In our research we build machine learning models to predict the deny and allow preferences of smart home data collection to identify circumstances that can lead consumers to change their preference and to predict the dollar value that consumers are willing to pay or get discounted in exchange for their privacy in smart home devices.

So given this goal, we started our research with a survey to collect consumers’ preferences. We have a detailed description of the survey protocol in the paper. If you're interested, I encourage you to read. Here I'm only going to present some high level overviews. So the survey consists of asking users for comfort levels, allow and deny choices, and notification frequencies, for four randomly generated information flows with combination of attributes, purposes, and devices.

Here's an example of such a combination. So in this particular case, the manufacturer or developers of smart home devices is accessing or inferring indoor location. They use this information for user tracking and profiling. And then we ask them their feelings about data collection, their allow and deny choices, and so on.

And then we provide a scenario based economic question to understand the money dollar value people put on their privacy. In this scenario, we assume the voice assistant costs $49, and we ask them how much they would be willing to pay extra for added privacy protection, or take as a discount or refund in the purchase to allow the manufacturer to collect and share their data.

The image on the left shows the distribution of dollar amounts given in each economic scenario as a percentage of the $49 price of the assistant. The red bar represent what the users are willing to pay extra to protect their privacy if their privacy features are not included in the devices. The green bar indicates what users are willing to take as a refund or discount in order to give away their data if a privacy feature has been included in the product.

You can clearly see that the average of both green bars are higher than that of the red bars. The key takeaway message here is that our participants are generally willing to pay less for extra as opposed to take a discount to give away their privacy. The image on the right shows what situation can make the consumers feel more or less comfortable. For example, factors such as consent not given, data not sensitive, and data usage are beyond primary use are the top reasons that can make people less comfortable; whereas factors such as having consent, users can control data, and not sensitive data collected are the top reasons why people feel more comfortable.

So this perspective enables the model to capture what contextual factors when present or absent can change people's preference of comfort level towards an information flow. Using data we collected through the survey, we build machine learning models too, which we're going to present next. So we build the model with PySpark which can be used on large scale, and sklearn because it's accessible and easy.

So again, we have more details of how we construct the model in the paper. Here I'm only going to present one example. So we hope that all developers can use this model to derive actionable steps for a large number of scenarios. In this example, scenarios were created using numerical features of the average user and all combinations of attributes, purposes of use, and devices in our study, and that ends up was roughly half a million scenarios for an average user.

In the majority of scenarios, as you can see here, it would make the user more comfortable if they could control the information flows. On the other hand, in the large number of scenarios, an average user would be less comfortable if the data are not used only for primary purposes. In this particular case, a large number of scenarios indicate that energy used for targeted ads would make the user more comfortable if they can benefit from it. That same number is actually greater than the number of the scenarios that would make the user more comfortable if they were aware of the fact that the energy use is being used for target ads.

So for the average user, that means the prediction shows the benefits weighs more than awareness. Lastly, in general, for an average user in the majority of the scenarios, it would make them more comfortable if the information is used for the purpose of home safety.

So what does it mean to the developers? These provide actionable steps for developers to implement appropriate information flow for their user. It is worth noting that a fine grained prediction is also possible if, for example, narrowing down the different attributes, purposes and devices. This model can also be used to understand how much value the user puts on their privacy.

In this table, the left column represents if the privacy is not included, how much money an average user is willing to pay extra before and after the purchase in order to protect their privacy. And the right columns represents if the privacy features are included in the devices, how much an average user is willing to get discounted before the purchase or get refunded after the purchase in order to allow developers to collect their data.

According to our model, an average user of our dataset is willing to accept $44 as a refund after purchasing the device in order to be more liberal about sharing their data, whereas they're willing only to pay $31 extra before they purchase the device in order to add more privacy features to protect the same data. The more concerning figure of this table is this one. It means that if the device does not have privacy features included, the users would only pay $28 after having purchased the device in order to protect their data, which is the lowest number in this prediction. That suggests that if the home becomes a place where privacy is not included, it will be difficult to make consumers pay for more privacy, which is great news for abusive developers.

So to summarize, from our survey, it is clear that secondary uses are not OK, but consent control and awareness can greatly affect users preferences. Second, consumers value privacy more when they have it than when they don't, but they already expect privacy to be included when paying for the device. Third, if the smart home adoption will grow regardless, it's our social responsibility as developers, researchers and regulators, to ensure that there are means to protect the contextual integrity of the home.

Other takeaway message from our study is that our models can be used by smart home developers to engage privacy in a practical way, deriving actionable steps and understanding the value of privacy of their users. Thank you. This paper has just been accepted by PETs--it's available through this link--and also with thanks to funding agencies of all the collaborators. Thank you.

[APPLAUSE]

YAN LAU: Thank you, Yaxing. We'll now turn to questions. We'll start with some observations, but I encourage you all in the physical audience to raise your hands and obtain a comment card to submit your questions. And also for those watching out in the webcast, please tweet us your questions to @FTC using the #privacycon19.

All right. So I think the presentations that we've heard for this session have sort of got a set of basic questions in a slightly different way. So like what do consumers think about privacy? What are their expectations and hopes about collection and use? How much control do they have? What affects their understanding and their willingness to trade privacy consciously or subconsciously for some benefits? Is it contextual?

So with that thinking in mind, I'd like to turn to sort of a specific set of questions for certain panelists, but also open to everyone. So this is for Katie, but also other panelists, feel free to chime in. So Katie, in your study you've documented interesting changes in consumer attitudes and understanding over the years. Where do you and also the other panelists think such views are headed in the next 10 to 20 years with each new generation of users, and what should such changes mean for public policy and regulation?

KATIE MCINNIS: That's a great question. So I advocate for better rules and stronger protections for consumers all the time. So I'm going to be a little biased and say that I think that wherever we are in the future in 10 to 20 years definitely depends on whether or not we have a federal general data privacy law or comprehensive state laws that cover these issues and allow consumers to have actual tools and controls over their data.

But I think that this sense of complacency that we're feeling, I feel like there must be some pushback that we're already seeing to the big platforms and a greater sense that this data is consumers’ and should not be used to undermine or to take away their autonomy. So I'm hoping to see better consumer understanding of their own personal data and hopefully a pushback and a greater awareness that they have the ability to say no in some instances, and others they should ask for laws to protect them.

ANDREA ARIAS: Any other thoughts on that question?

KRISTEN WALKER: Well, I worry that this issue is moving at such a speed that it's going to end up being much like manufacturing waste and the cleanup that was required after, much like climate issues, et cetera. That if we don't take action now--and I think Katie said that really well--then ultimately we're leading to long-term problems that are going to be harder to handle and deal with. And that will negatively impact consumers.

ANDREA ARIAS: We have a question from the audience, and this is for you Kristen. Can you discuss effective education efforts at any age? Because you divided it by different age groups. Should privacy oriented education be required, and at what age should it begin? So maybe take us through some ideas.

KRISTEN WALKER: Well, that's interesting because we look at children--and part of I think the focus on children isn't because we think adults understand this issue, it's because they are a protected class and it's an easier argument in terms of policy and regulation, to be quite honest. So I think educational campaigns are necessary, and they're necessary both in the media. And that would involve maybe a national campaign among different age groups that was most importantly targeted and tested. I think the Real Cost Campaign did that and did that well. But also included in educational settings. So I have some other work on ed tech, and that's concerning to me as well.

In terms of the ages, I'm not really sure because I see kids on their phone younger and younger and younger. And if it's not on their own device, it's on their parents device. So I would say as soon as possible.

YAN LAU: I know other panelists have done online surveys, and in developing these surveys you might have focus grouped or change certain aspects of it. But do you think you could address this question about educating not just children, but other age groups as well?

NOAH APTHORPE: Yeah, sure. So one thing I think that we have noticed while doing these surveys is the amount of confusion that there is about these issues from users. And we heard about this in some of the projects in the session and in the previous session. And what that really sort of says to me is that privacy by default I really feel like is important here, because to some extent, when we're asking about privacy policies or we're asking about any of the surveys that we did or that the other panelists did, these are complicated technical questions that researchers-- that manufacturers don't have the best answers to yet. And then to go and expect that average users should be the ones to take responsibility for their privacy seems like a big ask. And I would say that really, it's the responsibility of the developers, of the regulators, of the researchers to make the settings of these devices and these products preserve the privacy of the users without them needing to dive into the details.

And sort of just like when you are looking at options for any other products, say it's your car, we don't expect people to have a deep technical knowledge of how these things work. We expect the sellers, and we expect the dealers to be able to make decisions for them. And I think that really, it's going to be important in this space even more so because of how rapidly things are changing.

And I also think it's important for the regulators to realize that it's the consumers here that need the protection, and it's not necessarily irresponsible to go and make strong statements about data collection, because the companies and the manufacturers will be able to create the products that they would like. And I'd really like to see more-- we'd like to see, and we think that the people in our surveys and our focus groups and interviews we've done would sort of really like and often expect, even though it's not true, for their privacy to be protected by default.

YAXING YAO: Another thing I want to add is from our price study, when we try to understand how people understand the working mechanisms of targeted ads. So we found that people have all the different mental models of how technology works essentially. So I really believe in the power of education, educating a user of what they can do. Because a lot of people, they just feel like-- from our study, we observed, they just feel like they're helpless with all the ubiquity of devices, sensors, and everything.

But since they have different mental models of how things work-- and as Noah just mentioned, it's already a lot of responsibility for the users to-- we have put on the user. So I really think that education should be tailored towards users' ability and their age groups. For example, when you talk to-- it should be more personalized, right? So trying to understand how they believe technology works and how do we design either tools or how do we implement regulations to reduce the burden we put on the users so that they can better absorb the education we were trying to provide and to better protect themselves.

ANDREA ARIAS: So I'm glad that you mentioned regulation, because we have a question from the audience that I think hits on this. And it's focused on COPPA, but I think we can think about it just generally about any privacy regulations that we have. So it was mentioned that COPPA's language is vague, right?

So what can lawmakers or regulators do better to ensure compliance while not creating a paradigm that cannot respond or grow with technology? So if we are too prescriptive in our rules, then it's been bound by those rules without allowing technology to evolve as rapidly as it does. So what do you all think about what lawmakers can do to ensure compliance and obviously to maybe align with consumers' expectations on privacy?

NOAH APTHORPE: Go ahead.

ANDREA ARIAS: I open it to anyone.

KRISTEN WALKER: If I can take that for a second, I think what was important in our research was that the cognitive defense strategies that we used sort of was all around digital literacy. And in particular, we sort of talked about information sharing in that educational video clip and in the quiz. And so that's data literacy.

And those kinds of nudges and campaigns, as I noted in the implications, I guess, of our study, I think those are really useful and something that can be done. You know, industry really would be OK if we told them to do it. They want to do something.

KATIE MCINNIS: And I'll just push back a little bit on COPPA being too vague. We often advocate for other groups that test products for privacy and security in the children's market because it's just such low-lying fruit. They're not even getting verified parental consent before they collect data. They often include ads that are directed at children and ask them to make purchases in the apps.

There's been a lot of research over the past year showing that even though COPPA itself may not have a lot of rules, a lot of companies and apps and services are not even meeting the bare minimum of what COPPA requires.

NOAH APTHORPE: Yeah, and also I think that one of the most important things that regulators can do is make sure that technologists are involved in the process. I think often we see, especially in draft regulation or in proposals, that there are sort of well-known procedures or things that could take place that are either industry standard or that are well understood in the academic community that aren't acknowledged. And bringing in researchers, bringing in consumer advocates early on and often during the process I think could make a big difference in making sure that the final result of the regulations has more of the intended effect or that doesn't have technical loopholes that would have been caught by someone who is more familiar with the technical details.

YAN LAU: So we have a question from the audience, maybe for Mahmood or Yaxing. So as far as privacy valuation is concerned, do you know of any companies that are currently thinking about putting a price on data, quote? Will this be more widely adopted? If so, what does the timeline look like?

MAHMOOD SHARIF: So I'm familiar that there are some companies that I think are in early stages who are trying to make some sort of an information marketplace like the one that I described. Now, it's so early that I'm actually not familiar with the details and how it's going to work exactly. But it seems like an interesting concept where people could actually at least monetize their moneyif it's going to be monetized by others as well.

I guess similarly, there are some ideas where you could actually pay for-- use a browser that prevents tracking and also ads as well. But in return, you'll have to pay some amount, some subscription amount like $10 a year or a month so that this would be prevented. And the money would go to the service providers.

YAXING YAO: Well, actually I was thinking, when I was still a master's student-- that was like four or five years ago. That was an app that was in the Seattle area. There was an app. The goal of the company is to provide other companies who, if they want to open a store or things like that, they want to know the traffic of the people, of the populations so that it can pick a better location.

And then they created an app. They ask you to share your location every week, just one time every week for a whole year. And every time when you share a location, they pay you a certain amount of money.

Ironically, as part of this research, I actually signed up for that. I got about $120 for the whole year. But what surprised me was that at the end, they sent an email to the all the people who signed up for that service and said basically they're going to discontinue this because they have collect enough data to proceed, and that was a huge success they said in the email. So that means that actually people-- a lot of people are willing to get some incentives in exchange for their privacy. Yeah.

KATIE MCINNIS: As a consumer advocate, I just have to push back on that concept that you should be monetizing your data. I think that would just further exacerbate the inequalities that we're seeing in our society and also would lead to a situation in which privacy is more of a luxury and not actually an inalienable human right as it currently is.

KRISTEN WALKER: Well, and I think for valuation, there is a long-term price, right? Future research.

YAXING YAO: One thing I wanted to add, I didn't mention it in the talk due to the time limit. One thing we found that in our study is in the survey, when we asked people about how much money they want to pay extra or how much money they want to get as a refund in exchange for their privacy, we have about 60% of the respondents put a zero dollar amount, meaning that they don't want to get a refund in exchange for their privacy. That's an indication of people think they shouldn't put a price on their privacy. You know, we have more details in the paper if you're interested, but that's something I feel like it's kind of interesting.

ANDREA ARIAS: So let's talk a little bit more about that pricing, and then open it to everybody. But obviously you said that the zero value, but I know that you talked about during your talk that often once you buy a product, especially when you buy a product, and you have it in the home, you're not willing to pay as much to bring privacy into the home, right? Which this in turn limits the economic incentive for manufacturers to include these protections in smart home devices, particularly once they're purchased. Once they're in the home, what's the incentive then for the developer to add more privacy and bring it into the home after the fact? So I'm curious as to how you all think that we can overcome this problem of incentivizing these manufacturers, particularly once the products are in the home, to bring in more privacy protections for consumers.

NOAH APTHORPE: OK, yeah. I really want to echo Katie's point here is that this question in conversation is concerning to me, because it feels like we're operating under the situation where not protecting the privacy and collecting information is the norm. And then we talk about different purchasing strategies to support, to add privacy afterward. But it seems like instead what we really should be assuming or advocating for is a situation where the privacy preserving device or the data collection doesn't happen by default.

And I think that since we're not in that situation, this is a time where regulation really could have a strong effect. I mean, we've seen in the past with things like food safety regulations. That was not popular by the food production companies at a point, but it was obviously, in retrospect, a good idea to protect consumers.

And I think we may be at a similar situation here in terms of privacy and data collection where it's the role of consumer advocates and of regulators to really say, look, in this space the norm is that consumers' privacy will be protected and that data won't be collected. And that's somewhere where I think we have an opportunity to kind of lead from the front here rather than from the back to try to catch up with what companies want to do.

KATIE MCINNIS: In addition, maybe people aren't willing to pay for privacy protective tools in the home, but it's also we have a very new marketplace for those kinds of issues. I think the market's only going to increase, and groups like Consumer Reports are testing products for privacy and security. So we're essentially trying to make sure that the seat belt's already in the car before you buy it. The privacy protections should be built in before you buy the product.

In addition, those companies should have the right incentives to protect that information once they collect it. And those incentives are not available now. So these kinds of products should not-- the choice almost is not fair to the consumer at all. They're also considering price, accessibility and in many cases lock in with the company that they've already kind of bought some of their connected devices with. So I'm not sure if the setup here in the marketplace is truly representative of what consumers want and need. And in some cases, we should be regulating the market so they don't even have to think about it.

MAHMOOD SHARIF: I think ideally regulations should incentivize companies to bake in privacy protections by design. And also I think that we can see already that certain companies are realizing that by protecting their customers' privacy and respecting it, they can help bring new customers or even retain the existing ones.

YAN LAU: So this is for Noah and Kristen but also other panelists. So given the research into children and parents you've presented today, do you have any advice to parents on what they can do or how they can communicate with their children to protect themselves online? Noah, given your survey of parents, do you think there's anything that can change about their own expectations to help their kids? And, Kristen, from your experimental results, what is the best way for parents to educate children?

ANDREA ARIAS: Especially since you found that it works.

[LAUGHTER]

NOAH APTHORPE: Well, I think that one thing that stood out to me-- and I also want to reference here some previous work that we've done looking at actual toy implementations. And we found some really terrible practices in place where toys were sending location data of children, genders and ages when they had a crash or when they had a problem to third parties. And so first I think for parents it's important, one, to be skeptical, because even when toys are posting privacy policies or making claims, those aren't always backed up by what the toy is doing in practice.

And then, second, I think for educating parents, as you said, really kind of here we're running up against the same issues that were talked about in depth in the first session where there are various ways privacy policies with their issues are in place, but none of those are really working well. And additionally, often the settings options which toys and I think other electronic devices are providing are often opaque and don't necessarily provide meaningful choice.

So this is something that's very much an open question about how best to do this. Although I think from the work that I presented today, there is promise in the fact that toys and other products can be regulated or developed or designed in ways that do preserve and line up with expectation. So there is a route forward. The question is really what's the way to get there.

KRISTEN WALKER: So in our study, we showed that the perception of parents' restrictions sort of influenced behaviors and attitudes. However, other research that I do really shows how much parents don't understand or know, and I think that worries me as a parent in the sense that the reliance on apps or the reliance on other devices to protect your children sort of makes this a much messier context, I think, as we've kind of heard in all the presentations today and complicates that somewhat.

The good news I think is that we're starting to talk about this more. So you see New York Times talked about privacy for a month. They're continuing that. So as soon as this bleeds into the media, I think it will get more attention. I flashed-- as Noah was talking, I'm sure I see a Sesame Street episode in the future. So, you know, those kinds of things are where it starts.

And I think, and I can say that my co-authors think, that a national campaign would be really useful for driving that discussion. And it's something that doesn't influence market forces negatively. It doesn't require a lot of effort by policy or regulators in terms of money. And it makes demands of industry that are quite honestly long overdue.

ANDREA ARIAS: So I'm glad you mentioned the national campaign, because I'm curious as to what you think this national campaign could look like. Not only even for students, but since you mentioned that even parents maybe don't quite have a good understanding, could it be extended to other demographics as well, right, and into the national populace as a whole?

KRISTEN WALKER: Yeah. And I think that's important. The Real Cost campaign is a good example, as I mentioned. But it really does need to be something that is maybe funded or put together by industry in conjunction with other groups but also targeted and tested, right? It has to work. And if it's working, then great. We can continue doing that. But we certainly don't want it to not work, especially with children.

And the good news is that when you do this kind of outreach with parents, it has, you know, downstream effects. So that's a good thing.

YAXING YAO: One thing, so-- so far we've been talking about this. One thing I've noticed is that we have been focusing on-- regardless of children or average user, we're talking about protecting their privacy when they are-- as the primary user of the devices or whatever. But as these devices are getting more ubiquitous, like, when you go to your friend's place, are you going to-- if your kids are playing with his friend's smart toy with a camera with those kind of sensors, how are you going to deal with the privacy, your kid's privacy, in that situation? As you are not the owner, or you are not the primary user of the device, but you are still facing the potential data collection of your own kids. I think that's probably another aspect that can be considered and can be investigated in future research.

KRISTEN WALKER: Yeah, their friends' smart homes, right? Yeah.

ANDREA ARIAS: Yeah, so we'll do one more question, and then I think we'll break. Noah, obviously your research focused on COPPA, but I think you mentioned that it could potentially be extended to other frameworks. You mentioned HIPAA during your presentation. Do you think you'd find similar results in HIPAA or other regulations that in fact those regulations align with the expectations of consumers?

NOAH APTHORPE: Yeah, so I obviously hate to make too many predictions before the research is done. But I think that the general finding, which is that regulation does have the ability to place stipulations on data collection to force them to be more in line with consumers' expectations, I expect that that would still hold. And I think that in the HIPAA context, this is a context where there's sort of very strong existing social norms.

So we sort of have a way to think about when you go to your doctor, the information that you share with your doctor or with your nurses or with your hospital is very different perhaps from the information you might share with your employer. And that relationship is already well baked in to the culture. And that means that it's something that regulation can latch on to and extend into new contexts like medical devices.

But I think it's also important to point out here that when we're talking about HIPAA, a lot of the applications for the home IoT team medical products aren't necessarily covered by HIPAA. And that's something that I think a lot of consumers may not be aware of and also something which is problematic, because it means that those existing regulations can't be brought directly to bear against this new context.

YAN LAU: So I think this is sort of all we have time for today. But I'd like to thank our panelists for this wonderful discussion on consumer expectations and behaviors. If we could give a round of applause.

[APPLAUSE]

So with that, let's break for lunch. So please be back in the auditorium shortly before 1:40 when our next session starts. Thank you.