FTC PrivacyCon 2017
January 12, 2017
Segment 3
Transcript

SPEAKER 1: We're going to get started with session 3 on consumer privacy expectations. We have another great set of panelists here. And we're going to start with Jens Grossklags and Yu Pu who will be presenting their joint work.

JENS GROSSKLAGS: Thank you very much, and welcome to the last session before lunch. My name is Jens Grossklags, and this presentation is joint work with my co-author Yu Pu who will present the result section of our talk. The theme of our talk is predominantly focused on while not perhaps another trend, but something that been becomes more predominant that is that individuals are put in the position not only make an adoption decision for applications and to share their own data, but in the process of that, often they also share data about their friends, their family members, and acquaintances. And frequently these sharing decisions and also involve very personal and sensitive data about these affected individuals.

We refer to this phenomenon as interdependent privacy. To address this problem area, we are asking a number of questions. First and foremost, to which degree do users who are put in place and who are making these sharing decisions care about their friends' and their acquaintances' privacy? Or to put it differently, are we good stewards of the data of our friends when we are put in place in these sharing decisions?

As I said, there are many scenarios where this kind of sharing decision can happen. And particular, mobile computing nowadays, we are looking at the context of social third party applications where often certain permissions are referring to this the videos.

So let me illustrate what we are talking about on this schematic graph. So we have one user who is put in place and is faced with a decision to adopt an app or not to adopt an app. Often then there is some permissions screen shown to the user. In this particular case, it involves some permissions affecting the user himself or herself, but also permissions affecting all the users that this individual is connected to. In particular photos and the birthday information.

Now when the user decides to adopt the app, then of hundreds or even thousands of users that the user is connected with, their information will be also transferred over to the third party. And there's an important distinction. Why the direct path from the initiating user to a third party application can be considered as direct decision making, often the other individuals have very limited opportunities to do something about the state of sharing.

Now moving on, what this our approach? And what we're doing is a combination of different methodologies. First, we use an approach that allows us to quantify how much individuals care about other's data by conducting a so-called conjoined analysis.

Here users are asked to make trade-off decisions over different versions of social apps which differ regarding the data collection practices. On the one hand, regarding what information is

collected about the individual. So personal privacy, but also what is collected about all these other individual sets of users connected to so-called inter-dependent privacy.

We also consider other attributes that are important in the app adoption context. So we compliment this conjoined analysis with a second methodology by collecting survey data to explain the evaluations that we solicited with the conjoined analysis. And evaluations and the survey responses and together enable us to build a comprehensive behavioral model using structural equations modeling. So in addition to that fee, consider two key factors, which are relevant in the scenario of social app adoption.

On the one hand, sharing anonymity, on the other hand, context relevance. So first we explore how an individual's valuation of other state that differs based on whether the app adoption decision and associated data sharing with a third party is identifiable or not. In addition-- and this was a common topic in the previous sessions-- we also look at whether data collection is actually [INAUDIBLE] from them for the purpose of the application or exactly the opposite.

And now for the explanation of the results, I hand over to Yu.

YU PU: So we calculate the value of our interdependent privacy across three dimensions. That is of friends basic information, friends valuable information, and friends full profile information, which is sum of both friends basic information the valuable information. And we have detected the same effect for sharing anonymity and a context relevance across these three dimensions. And to take a friend's full profile as an example, when we compare the two black bars and the two grey bars we can know that when individuals know that sharing friend's data is anonymous, they would be likely to put less valuations on their friend's data.

When it comes to the effect of context relevance, we also find that when individuals know the information of friends is useful for apps' functionality, they would also likely to put less valuation on friend's data. This indicated that individuals might trade off friend's data for better app performance.

So note here when we talk about value of interdependent privacy, we are referring to the value of the information of users of friends. However, consider that a user has on average of around 300 friends, we further break down this valuation. That is, we also calculate the value a user place on a profile of a single friend. And such valuations are represented by the various small tiny black bars there. And to take full profile information as example, while individually put around $2.5 on his own full profiling information, so when it comes to the full profile information of a single friend, such value drops to as low as $0.07.

Consider that the profile information of a friend includes also sensitive information such as friend's birthday information and a friend's photos and so forth, this indicates that individuals can be considered as privacy [INAUDIBLE].

We next explain the factors that affect the interdependent privacy valuation, and we further break down this model into three pieces. Let's first look at the top of the model where we verify the existing models that explain the factors that drive the concern towards own privacy. And similar

to the existing work, we found that factors such as disposition to value privacy, as well as perceived control of personal data affects how people concerned about their own privacy.

So what is novel in our research is that we also explain the factors that drives people's concern towards other's data. In specifically we found that when individuals have higher privacy knowledge, is they would be more likely to place higher value on friends' data. In addition, we also found that the degree to which people care about others, which is measured by other regarding preference here, also affects how people concerned about others privacy.

In addition, we also investigate the relationship between privacy concern and the privacy valuation. Particularly, we found that when individuals have high concern towards friend's data, they also put higher values on friend's data. And for the treatment effects. we found that the same effects as we discussed previously.

JENS GROSSKLAGS: So our work has numerous privacy policy implications. First, we find and identify numerous individuals, but also a range of broad scenarios where the data of a single friend is valued-- oftentimes, it's a mere fraction of a single cent. And that really raises the question of to which degree we should limit the sharing of friend's information on mobile networks, social networking platforms, and so on.

Question that are arising are what interventions are actually suitable, and whether social networking platforms but also app developers would be willing to self-regulate in that regard. We also find that awareness and education can have a positive impact, but it's only one factor as our complex behavior model shows. So multiple factors need to be taking into consideration to elevate perceived value of inter-dependent privacy.

Finally, our work has also a number of implications for the design of sharing decisions. And on the one hand, we can observe that users are willing to trade off the privacy of those that they're connected with in favor of having better app performance. On the other hand, they are also willing to discriminate and to punish providers that actually solicit oral event information with respect to the fulfillment of the app's purpose. We unfortunately also observed that under the veil of anonymity, users are more often willing to share others information and value the information of others less. So both of these problems can potentially be addressed by transparency enhancing mechanisms, which were partly also discussed in the previous section.

So the final slide that we're having is just an overview of our working papers. Today's talk is mostly based on our current working paper. The other papers however, serve as an important complement and validation of our findings, which gives us great confidence in the robustness of our work. Thank you very much.

SPEAKER 1: So our next speaker is Chanda Phelan from the University of Michigan.

CHANDA PHELAN: Does this picture creep you out? Would you risk having these crawl on you in exchange for say, $500? What if I told you that these guys are called tailless whip scorpions are totally harmless?

Would that make you any more willing to take the dare? I would. It's still really creepy though.

So that's the takeaway. People can be uncomfortable, think that something is creepy, even though when they think about it, they're rationally OK with it. And sometimes people still feel creeped out, even though they've decided they shouldn't be bothered.

Same thing happens with privacy concerns. I'll argue that there are two components of privacy concern that are related, but ultimately independent. Someone can have either kind of concern without the other.

These two types of privacy concern correspond to two kinds of human information processing. You might have heard of them described as system one and system two, if you've read Daniel Kahneman's book Thinking Fast and Slow. System one is the intuitive process. That's the one that generates impressions and works off associations and heuristics. It's automatic, fast, often tends to have an emotional component to it.

System two on the other hand, is the reasoning process. It's the one that generates judgments. It's under conscious control. It's effortful and slower, but in exchange, can be governed by logic.

So to sort of criminally overgeneralize it, system one is your lizard brain or your autopilot, and system two is your experience of consciousness. And because system one is so much more efficient than system two, it's system one that's usually running the show and only tosses things up to system two when it runs into a problem that it can't deal with itself. And system two is still going to be heavily influenced by the impressions of system one, even when it's fully engaged.

So in our study, we talked to a bunch of undergrads about their privacy concerns surrounding privacy invasive tracking extension for a research project called M Together that they'd chosen to install on their browser. In talking to them, we noticed some patterns about their kinds of concerns that they expressed. Some privacy concern seemed to come from system one. They were emotional, immediate, and sometimes couldn't be articulated very well. The other ones were from system two, which is a rational weighing of risks and benefits.

Here's an example of intuitive concern from an interviewer. The interviewer asks would it change how you felt about M Together if it read your messages. And subject five knows immediately, oh definitely. That's pretty invasive.

Interviewer asks what do you think is different? And subject five has no idea. He tries to explain, but then just sort of end up finishing with I don't know. It just kind of makes me less comfortable.

The first response from subject five is the fast gut feeling. He knows immediately how he feels about the intrusion, but once prompted by the interviewer, has trouble articulating the reasons for it. The second type would be a more deliberate assessment and doesn't always happen. So here, subject five doesn't move on to the second stage of consideration until prompted by the interviewer, and doesn't really complete that assessment of considered concern.

In the rest of this talk, I'm going to build up to this conceptual model of how people decide whether or not they have privacy concerns. People run into a disclosure decision, and right away they get a gut feeling. The assessment of their intuitive concern.

If it's not creepy, then they don't think about it anymore. They just move on to their disclosure decision. If it is creepy, then that's when they assess considered concern, at which point they decide whether or not they're bothered by the intrusion.

In our interviews we found three factors that most commonly affected our participants' assessment of privacy concerns which were social presence, low marginal risk, and trust. First off, I'm going to cover social presence. The sensation that someone is looking over your shoulder as you're browsing.

We found that our participants mentioned social presence when talking about high intuitive concern, and made no mention of it when talking about considered concern. So participants who felt social presence described data collectors like they were people who had the participant under surveillance. Subject 27 is talking about somebody behind me trailing me. And subject four is talking about someone sort of watching you.

The concerns here are intuitive. We've got these emotional reactions like subject 27, it's just a little scary, or subject four who says it's just a weird thing to think about that someone's sort of watching you. And with subject four, you can also see that he's having trouble articulating the reasons for that feeling. He knows what he feels, he just doesn't know why.

Mapping this onto our model, we theorize that social presence increases intuitive concern. Someone may decide they're not bothered by an intrusion, but the intuitive feeling of creepiness tends to remain, even though it doesn't affect they're considered concern. The next factor, impacting privacy concern, is a low marginal risk. So here meaning an assessment of additional risk from some disclosure compared to other information that's already been disclosed.

An example of this is subject 30. All you guys were asking for was monitoring my sites and my hits, and basically a lot of other sites already do that without my permission. The impacts of this marginal risk assessment seem to be the inverse of social presence.

Our participants use low marginal risk to justify low considered concern. And it didn't appear to change intuitive concern. And you can see that in how subject 30 says all you guys were asking for.

Or subject 11 who says it did occur to me like, oh, what if they can see my Facebook. But in the end I just signed up for it. So even though he says he's not bothered, we can see that subject 11 isn't particularly happy about the situation. And that's because low considered concern won't necessarily erase that intuitive feeling of creepiness. His system two has overruled his intuitive concern, but the intuitive concern is still there.

Finally, trust. Specifically trust that the institution or individual will use your data properly. We found that trust was associated with both lower intuitive and considered concern. In the context

of M Together, it largely eliminated privacy concern, even when people said they would have been concerned in other situations. But to an extent, participants trust in U of M is rational, but it could go beyond rationality.

A lot of our participants were unaware that M Together was gathering any of their data, which suggests that the extension never registered as creepy. So they didn't think about it. Never engaged in considered concern.

Others were a little more cautious. As you can see in subject 8 who says I was just flipping through. Yay. Whatever, install. And then when I went and looked back, I was like, wow. They must be collecting something in my computer.

So I guess I was maybe hesitant. I feel like that's not their motive to collect personal information from me, especially when it's coming from professors from the University. They're trustworthy people.

So here, he does get that gut feeling of creepiness once he knows what the extension is, but uses trust during his assessment of the intrusion is not problematic. And when we map that onto our conceptual model, he hits a red flag in intuitive concern, thinks it's a little bit creepy. But once he thinks about it, he decides he's not bothered.

So when we put this all together in the full conceptual model, we have the properties of intuitive concern mapping generally to system one, the properties of considered concern mapping generally to system two. And social presence, low marginal risk, and trust all having different impacts on the two types of concern. So I want to leave you with two implications of this conceptual model.

One is the new explanation of the privacy paradox. The privacy paradox describes how people tend to say they have really high levels of privacy concern, but then give away their information in return for very little benefit. One explanation of the privacy paradox is that people act on surface cues that may or may not be relevant.

So for example, people tend to disclose less when you reassure them of how careful you're going to be with their data, because you've drawn attention to the fact that their privacy is at risk, they engage system two to assess considered concern. Whereas without the warning, they never would have gotten around to it. The conceptual model summarizes this nicely, but it also gives us a second explanation. Someone can rationally decide that a disclosure is OK, that there is no considered concern, and yet still feel creeped out. Like the participants who decided there was no marginal risk in extra disclosure, but still felt uncomfortable if it felt like there was a person looking over their shoulder.

The second implication goes more directly to policy. To the extent that policies are supposed to protect consumers from being tricked into doing things that are against their interests, we should define their interests primarily in terms of considered concern, or what people would decide for themselves if they engaged in careful fact finding and reasoning about the costs and benefits of the disclosure decision. That means first that we have to be careful when we elicit people's

concerns. That we separate out the expressions of intuitive creepiness. Just because they say they're concerned doesn't necessarily mean that privacy advocates should be concerned unless it's considered concern.

Rather, I think privacy policies should encourage congruence. Low considered concern? Don't require lots of privacy disclosures or things that are going to make consumers feel uncomfortable. Conversely, if there's high considered concern, then there should be high intuitive concern too. Make people feel creeped out when there really is a danger to them so that they reason about it and protect themselves.

Our conceptual model is saying that it's possible for system one and system two to come to different conclusions about privacy concerns. In other words, it's creepy, but it doesn't bother me. Policies should be put in place to reduce that divergence. When it's creepy, it should bother me. And when it would bother me, it should be creepy. Thank you.

[APPLAUSE]

SPEAKER 1: Our next speaker is Yang Wang from Syracuse University.

YANG WANG: All right. Thank you for having me. So I'm very glad to talk about our recent research on people's understandings of how online behavioral advertising works.

And this is an important question to answer because oftentimes people's understanding on things would affect their behavior. So for example, your mental model where your understanding of how that door works will potentially influence how you open or close the door. So that's what we're looking at in this study.

So I'm sure that most of you in this audience knows what's online behavioral advertising, or OBA, but just in case, it refers to the practice of tracking an individual's online activities, in order to provide customized or target ads that will fit to this person's interest. So if you look at the previous literature in this area, a lot of studies have shown that people have mixed feelings about OBA. So on one hand, sometimes they find these target ads are useful and smart, but they're also concerned about these things could be very creepy or even very scary.

One of the limitations of these prior work is that researchers would explain what OBA is and how it works before they ask people's opinion about OBA. So we really missed an opportunity to understand what are people's own understanding of how OBA works. And this is the focus of our study.

So we drew from the literature on folk model. So I don't really have time to go into the details, but basically you can think about folk models as models of the reality that we have in our head to help us make decisions. And one of the important properties of folk model is that it can be incorrect representation of the reality. But nevertheless, people use them to make decisions in practice.

So why do folk models matter in this context? Well first, have a better understanding of people's mental models or folk models will help us understand why users have particular attitude towards OBA. And second, as I will show you soon, that a lot of our participants actually have incorrect or inaccurate understanding of how OBA works.

We're at FTC. And one of the task of FTC is user education. If you think about user education, it would be more effective if you can target at these specific inaccurate understanding. And last, as I hinted to earlier, these understanding can affect people's behavior.

So in terms of methodology, we did two rounds of in-depth interviews, mostly focused on the question of what are people's understanding of how OBA works. We asked some other questions, which are not really the focus of this talk. We did this with 21 participants in the west coast and the east coast. They had a pretty diverse demographic and occupational backgrounds.

So the way we get at people's understanding of how OBA works is by providing them a hypothetical scenario, which is supposed to be pretty common. So you first go to Amazon and you look for some shoes. And a few hours later, you visit Facebook and you found some other shoe ads on Facebook.

So what we usually do is we ask people's opinion about this practice, about this scenario. But what we did instead this time is in addition to ask them to explain what's going on, we also asked them to draw on a piece of paper about hey, what do you think happened in this scenario. So I'm going to show you the four folk models we identify, and some of the drawings from our participants.

So this is the first model, which we call the browser pool model. So basically people in this model believe that it is the browser that does everything. So the browser tracks their information, what they did on Amazon, store their information in the browser.

And then when the user goes to Facebook, the browser is going to retrieve the relevant ads and display it. So again, everything is done by the browser. So that's the browser pool model.

The next model is what we call first party pool model. This is very similar to the previous model, except that when user goes to Facebook instead of the web browser pulls the relevant ads, this time is the first party, in this case Facebook, will pull the relevant ads. The next model is what we call connected first party model. And this is a pretty interesting drawing. Basically, this participant was trying to show is that when she goes to Amazon, Amazon will essentially pass this user information, what shoes she's searched to Facebook.

And I don't know whether you can see this little dollar sign next to the Facebook box. So essentially suggesting hey, Amazon is basically selling this information to Facebook. And then, if you see the other side of the Facebook box, there are two dollar signs. Basically Facebook is now going to talk to other shoe ad companies and basically make a profit.

And if you look at the top of the figure, you see two straw mans. So basically the participant in saying that the reason that Amazon is passing these user information to Facebook is not because

the two CEOs of the companies are the best friends under the sun, it's just because there is monetary incentive. So that's the connected first party model, where again, Facebook and Amazon are directly connected through direct transactions.

And lastly, we have this model what we call third party model. And I actually give you two examples here. The examples on your left is a pretty simple example which basically this participant believed that there is some sort of gigantic internet space that's between Amazon and Facebook. And that's what happens. But they can't really pinpoint who represent the inner space and what they do and who's behind it.

The one on the right is slightly more complicated, but similarly, you see there's a big circle, which representing a mysterious database. Basically all the companies will contribute user information to this huge database, and then when they need to pull ads, they will talk to the database. But again, they don't know who's running this mysterious database.

After we talk with our participant, ask them to explain their figure, their drawings, and their understanding, we then explain actually, this is how OBA works. But this is a very simplified version of the OBA common practices where essentially a third party tracker will be placed on both sides so they know what you do and then they can build profiles, and then provide target ads. Of course we didn't really talk about the messy ecosystem about ad beating, ad exchange. So this is kind of the simple explanation we gave people.

And then the next thing we ask them is OK, so you have two things to consider. And one hand, we're talking about the information that are being collected or tracked. On the other hand, you are thinking about the trackers where the people who are collecting this information.

And then we're asking them OK, so which of the two is more important to you? Is it the information being collected? Or is it who is collecting this information? And 20 out of 21 participants all said that information is more important.

And this is a very interesting result because if you think about their current privacy tools for OBA-- the ghost tree app locks-- they are operating in a model that's based on trackers. So let's say you go to The New York Times, they're going to show you a list of trackers, and then users can decide to either reject either block or accept these trackers. But this result suggests that that's not really what people pay attention to. What they care about more is the kinds of information that's being collected. So we would argue that these two would potentially be more effective if the organized were a group that trackers based on the information they're collecting.

And just two more quick implications for privacy design and policy. One is that for those of you who are building tools to help or enable or empower ordinary consumers to make decisions or be aware of online behavioral advertising, your tools cannot assume that user know about third parties. As you recall, three out of the four mental models, people don't even know there is existence of third parties. In fact, that's 2/3 of our participants. They don't know there is actually third parties.

And another important implication is that the trackers right now we looked at a bunch of popular sites and the trackers and their privacy policy, many of them don't have a privacy policy. Even if they do, they do not say clearly what types of information they're collecting. So this makes it very difficult for ordinary consumers to make decisions. They don't know what information you're collecting about me. So they're suggesting is that either there should be an industry best practices, or there should be legislation that basically require trackers to say very clearly what types of information they're collecting or what types of information they're not collecting.

So I would end with that. And I thank my co-authors, Yaxing Yao, who is a doctoral student of mine, and [INAUDIBLE] who was a visit and researcher from the University of Rome. And if you're interested in more details about the mental models, please refer to our paper this year at CC. Thank you.

[APPLAUSE]

SPEAKER 1: All right. So we have one more paper in the session. And this is going to be presented by Mahmood Sharif from Carnegie Mellon University.

MAHMOOD SHARIF: Hello everyone. Today, I'm going to talk about a lab study exploring users in-context preferences for online tracking. This is joint work with my colleagues at Carnegie Mellon and Qualcomm.

I'm going to start by showing you a simple example of tracking with cookies. Cookies are those small tokens that can uniquely identify your web browser. They are used for many things including tracking.

The way they work is that when the use loads a website from web server, the web server might ask the web browser to store a certain piece of information. For example, that your ID is 1, 2, 3, 4. Later, when you load another page from the same domain, your web browser is going to send any cookie it has for that website or domain to the web server.

This enables the web server to identify you, and to serve you with custom content. For example, custom news articles based on what you looked at in the past. This is first party tracking.

Now I'm going to talk about third party tracking. Third party requests are those request that the user did not explicitly make. First parties might have content from many different domains.

For example, CNN might have ads from some advertiser. So when you load CNN, your web browser will send any cookie it has for that advertiser along with a request. This allows the advertiser to identify you and to serve you with targeted ads. This advertiser might be available on several first parties which enables it to learn more about you than any individual first party.

Experts have varied opinions regarding online tracking. Proponents say that online tracking allows targeted ads and customized content, something that both the industry and the users may find value in. They also say that the revenues that companies derive out of online tracking can enable them to provide free services to users. On the other hand, opponents say that online

tracking creates privacy concerns. That third parties can use it to build detail profiles about users, and that this can happen without users' knowledge.

We just saw what experts think. Now I'm going to talk about what users think. It's there it's important to understand users' preferences regarding online tracking in order to evaluate current tools and maybe provide and suggest better tools and regulation. Prior work has shown that the majority of users have privacy concerns regarding online tracking. And that users' preferences are complex, that they depend on the situation and about the benefits and the risks of online tracking.

However, most of prior work has been done with hypothetical situations, which is not ideal because users may not have enough context to make decisions. So for example, it's different to ask users how do you feel about online tracking on a shopping website, compared to how do you feel about online tracking when you are shopping for heartburn medicine on Thursday on Amazon.

In order to provide better tools, it's important to understand users' of preferences in the context of their own browsing history. And this includes understanding what harms and benefits they perceive, and what are the situational factors that affect users' comfort. Answering the first two questions is important in order to evaluate current tools and see if they satisfy users' needs. And maybe even suggest better tools and better regulation.

When I say that we studied users' preferences in the context of their own browsing history, what I mean by that is that our participants downloaded an add-on and used it to send us their own browsing history. They had the ability to go through their history and remove items they were uncomfortable with us seeing. Then we went through their browsing history and picked specific situations to ask them about. And then we conducted the interview.

So unlike prior work, we actually asked our participants about tracking that actually happened, and not completely hypothetical situations. We conducted 35 semi-structured interviews in which we asked about a variety of situations, including first and third party tracking. And eventually two coders developed a code book from a test set of interviews, and coded the entire set of interviews.

Here's an example of what we asked our participants about. So we would show them a sample from their own browsing history, and we would ask them what are the benefits of tracking, what are the harms of tracking, and whether they were comfortable with it or not. Now I'm going to show the results.

Our participants perceived the range of outcomes for online tracking. Some of these outcomes were overtly noticeable, while others were more hidden. Only noticeable after reflecting on them, or happening behind the scene unbeknown to users. Some of these outcomes were perceived as beneficial, while others were perceived as harmful. In addition to the outcomes that our participants perceived, we noticed that there were situational factors related to the nuances of different situations that affected their comfort.

Now I'm going to talk about both the outcomes and the situational factors. Here are some of the overt outcomes that our participants perceived. So most of our participants for example, had an opinion about targeted ads.

Many of them saw them as beneficial, because they showed them ads on products that they might be interested in. Others however, thought that they were harmful and annoying. You can read more about the overt outcomes in the paper.

Here are some of the hidden outcomes that our participants perceive. Notably, some of our participants thought that the revenues that companies derive out of online tracking were harmful because it made them feel used by those companies. Others thought that they were beneficial because it enabled companies to provide them with free services. You can read more about the hidden outcomes also in the paper.

We want to know in which situations our participants who are comfortable or uncomfortable with online tracking. And part of this are the outcomes of online tracking. Interestingly, in some situations, our participants could perceive the outcomes of online tracking to be harmful and beneficial. And while their perceived outcomes to be beneficial, it did not necessarily mean that they were comfortable with tracking and vice versa. So looking at the outcomes alone, was not enough to determine our participants' comfort.

And surprisingly, our participants were less comfortable when they [INAUDIBLE], and we also noticed that the hidden outcomes seemed to drive more discomfort out of participants than overt outcomes. So since the outcomes alone were not enough to determine our participants' comfort, we want to know what it was about specific page visits that made our participants more or less comfortable. We call these factors that affected our participants' comfort situational factors. Here are some of them.

For example, when visiting sensitive websites like banking websites, our participants seemed to be less comfortable with third party tracking compared to first party tracking. We also noticed that the kind of information being tracked affect our participants' comfort. So for example, some of the participants were comfortable with only general information about their visits being tracked, but they were less comfortable if identifiable information, like their names were being tracked. You can read more about the situational factors also in the paper.

We used our findings to evaluate current tools. And we looked at several tools including Adblock Plus, Ghostery, Private Browsing Mode, and more. To summarize our findings, we found that most of the tools adequately address the harms of online tracking. So for example, most of the tools can block ads that users find is annoying.

However, none of the tools that were looked at was able to allow the benefits of online tracking with small amount of configuration. So for example, none of the tools that were looked was able to allow targeted ads that users find that as useful with a small amount of configuration. We believe that is because none of the tools allow controls based on situational factors that matter to users.

We believe that with more detailed and precise understanding of users' preferences, it's possible to build better tools to control online tracking. And to that end, and this is very preliminary work, we want to see if machine learning can be used to build better tools. On a very high level, we use machine learning to predict whether participants are comfortable or uncomfortable for specific page visits based on the situational factors that we found matter to them. This way we can build tools that can block tracking if users are predicted to be uncomfortable, and allow it otherwise.

Now I'm going to present the prediction results. So on this graph, the x-axis shows the percentage of bad tracking allowed, and the y-axis shows the percentage of good tracking allowed. Ideally, we want to block all of the bad tracking and allow only the good tracking.

Here's how we do. So in the blue area, we do fairly well. We block the majority of bad tracking and we allow some of the good tracking. While this is nowhere near ideal, this is only a first look on how machine learning can help us in this space. We believe that with more data we can do much better and design better tools.

So to summarize, we explored the users' preferences in the content of their own browsing history, and found which outcomes matter to them, and which situational factors affected their comfort. We evaluate current tools and found that they don't adequately address users' needs. And we show that there is some hope for automated preference enforcement. Thank you.

[APPLAUSE]

SPEAKER 1: OK, we're going to have a brief discussion before lunch. If any of you have questions, please come to the microphone. Let me start with one, and then we will go to the microphone. So I think these are all very interesting papers. And there were a lot of common threads between them.

We heard about intuitive concern versus considered concern. And then we also heard about noticeable and hidden concerns that may require reflection. We heard about misconceptions that users have. So it seems that all of this suggests that without reflection, users don't always understand what's happening. And in some cases, even with reflection, they don't always understand what is actually happening.

So what does this tell us about the types of tools and policies that we need to address our users' expectations or concerns? You two give us what is your top suggestion here? Do you want to start, Mahmood?

MAHMOOD SHARIF: I think that part of how we can handle the situation is basically partially educate users. In order to have them at least have at least a small amount of information about online tracking and how it works. And then try to build better tools that can work when they don't have enough information. So build tools that can fill in the gap when they can't reflect.

YANG WANG: So I would say that the most important thing is that we should require trackers to tell us clearly what information you're collecting about us. I'll just leave it with that.

CHANDA PHALEN: I think that educating people on what exactly is happening is really important. Because at that disclosure decision, requiring a lot of reflection is going to backfire a lot. They won't even necessarily decide I'm bothered by this.

They'll just decide this is too hard of an decision. I'm going to think about it later. And then they never think about it. So either they'll be giving up the benefits of that privacy disclosure decision, or they end up doing something wildly wrong with their privacy. Something that would have bothered them if they'd been able to think about it.

JENS GROSSKLAGS: Attention of users is certainly one of the most scarce resources. So we have to support them in being able to do the decisions that actually matter. And I think what Chandra presented is one example of this broader space of bounded rationality research and the various kinds of burdens that we are facing in practice.

So we need to get rid of those kind of decisions that are impossible to do for users or are actually burdensome from an economic or psychological cost perspective. And identifying those is naturally very difficult, and may require also the involvement of technology. But the truth lead them to interventions, baseline regulation that frees us to focus on the things that really matter.

YU PU: So I want to say first of all education is quite important. Then I think of the mechanism and policies is also very important and to help individuals to make well-informed decisions. To apply to our case, we think it is very important to inform a user whether the sharing is anonymous, and or whether the data collected is important.

SPEAKER 1: All right. Let's take a question from the audience over here.

KRISTIN WALKER: Hi. I'm Kristin Walker. I'm a marketing professor from Cal State Northridge. And I love that you guys were talking about educating obviously. And I like that you're talking about decision making.

I wanted to ask a more macro question about expectations, because everybody wants information. I mean, it's pretty clear from what you say, consumers want information. Marketers obviously want information, policy makers want to make sure that that information is handled correctly.

But who do you think is right now shaping consumers' expectation about privacy? Is it consumers shaping their expectation? Is that technology shaping that expectation? Is that policymakers? Is it marketers?

Because, I think we obviously know that privacy is important. But making a decision about how to handle your privacy, is it after the fact, or before the fact? So kind of an antecedent question for some of your studies.

SPEAKER 1: If you'd like to take that.

YANG WANG: I guess I'll take that first shot. Great question. I think that your question about who is really shaping our influencing people's privacy expectations, I think is a confluence of many entities, stakeholders, certainly the markets where the industry has a pretty strong influence on us.

But I think increasingly, the news media, government, they also play a role. And also, not to mention the civil rights organizations that try to blow a whistle when these industries are doing things that are potentially privacy invasive. But I think moving forward, I would agree with my fellow panelist that I don't think technology or education or policy alone would be a silver bullet. You probably need a combination of all above.

JENS GROSSKLAGS: Looking, for example, at this legal standard of reasonable expectations with respect to privacy, there's definitely a conundrum emerging. So on the one hand we want to enhance awareness, you want to enhance understanding, but the standard also de facto tells us that once users know that particular technology is applied in the public domain, then we should not have a reasonable expectation with respect to privacy about. So this is kind of a vicious circle that we are facing when it comes to these kind of issues.

When it comes to tracking, if we increase awareness about tracking, does it mean that we have less of a reasonable expectation about preventing the most egregious facets of consumer tracking, and so on and so forth. So I don't have a perfect answer for that. But I think this is a circular process that is not often enough talked about.

CHANDA PHALEN: And I think you brought up a good point about considered intuitive Will more information really be helpful? So thank you guys.

SPEAKER 2: First of all great sense the Chanda for mentioning Daniel Kahneman. I wanted to explain that to behavioral economics and to the entire panel. There are some findings which could be used for the privacy. Most importantly, the prospect [INAUDIBLE] shows that people are about 2 and 1/2 times more adverse to losses than to gains. So the gains in usability probably should be balanced against two to three times losses and privacy.

In other words, if somebody could provide some quantitative model, what you gain in usability and what you lose in privacy, probably people would be more privacy conscious. How to make people conscious of that? Another behavioral economics find is availability bias. So tell people about some possible issues, give them some visible visual results, and they may become more privacy conscious just because they will become more alert, more scared, or more creeped out.

And to bring up another behavioral economist, Dan Ariely. He has a so-called IKEA model that if you build your own bookshelf, no matter how awkward it is, is you would value it. So if you make people work a little bit about configuration of their privacy, maybe that would cause them to value it more.

CHANDA PHALEN: One thing to add to what you're talking about with people being afraid of losses more than they like gains, we also have a problem in the privacy disclosure decision of

this certainty asymmetry, I guess you could call it. They have a very solid benefit that they're going to get.

Whatever they're signing up for, they want it. That's why they're signing up for it. They have this thing that they want right on the other side of giving up some of their privacy against this like maybe something bad will happen sometime in the future when the bad person gets-- There's so much there that when they have that carrot so close to them, I think they'll discount those potential losses a little bit more. Because they don't know what they are.

SPEAKER 2: This is a very dangerous comment, this is Washington, I would say that today a lot of people in the political parties are very, very concerned about their privacy. And maybe they were not as much concerned several months ago. And the point is that when you have a really bad example in front of your eyes, your psychology becomes also somewhat different.

JENS GROSSKLAGS: So clearly, behavioral economics and psychology is a great inspiration for all of us here on the panel. And we have used ti to various degrees in our own work. I think the art and science of privacy research is to try to find out which of these behavioral concepts truly matter in the context of privacy, but also how to translate them to the specific domain of privacy.

So let's, for example, use the endowment effect. So endowment effect has been shown for durable goods prominently. So the first question then arises, how do we translate it to information goods? However, information goods are not really privacy.

So because once privacy is revealed, once you've given away your information, it's kind of gone. So how do you position that in the context of the endowment effect? And I don't want to elaborate too long on that, but it's just one example of where I think a lot of work, a lot of careful experimental work, but also theoretical reasoning is needed to make these concepts truly useful and impactful in practice.

YANG WANG: So I would just quickly that there is a big privacy project called nudging, which was done by Alessandro, Lorrie, and Norman Sadeh. I was on the team. So we were looking at a building design to hope to nudge people towards making appropriate privacy decisions by mitigating the impact of these various biases.

SPEAKER 2: OK, thank you so much everybody.

SPEAKER 1: So to follow up on this, so one approach we've been talking about is nudging, but another approach is to have some automated decision making, preferably based on the user's expressed preferences. So what should the role of the automated decision making be as opposed to trying to educate and nudge users?

MAHMOOD SHARIF: I think that whenever we can take the burden off the decision from the user and make decisions automatically, we should be able to help them. At the same time, we should acknowledge that when we're making automated decisions, we might often fail because we don't have enough information. So we'd also need to monitor our models and see when they

don't have enough information. And when the confidence in the decisions are low, and try to pick safe defaults in those cases, like for example blocking online tracking, that case.

YANG WANG: Yeah, I think I agree with what you said. But I would caution that when you're making these automatic decisions, it can be very dangerous. Because if you think about the ethics of doing this, so why are we the researchers better positioned to make these decisions for the users.

You're taking part of their agency away. You can argue that they probably are not in the best position to make these decisions. But nevertheless, I think it's important to be cognizant about these agency issues.

The other thing too I want to say is that often times we built these automatic using machine learning algorithms to predict their future preferences, used on what they have done in the past. The assumption is that what they done in the past, they're real preferences that-- or in other words, these are things they will want to do. But whether that's the case, I think is a question mark.

JENS GROSSKLAGS: Well, that's certainly also the question then where is the technology coming from? Where are the automated mechanisms coming from? So seeing it from the user end, there's definitely then the problem of low adoption of privacy enhancing technologies more generally. So how could we improve on that by getting more useful technology in user's hands?

On the other hand, if you see it from the perspective of having some mandated technology, then the question clearly arises who is actually designing that? And who sets these mandatory standards? And what rules are folks up on the context of tracking embedded in these kinds of technologies?

So both are extremely difficult problems. The one is behavioral and availability problem. The other one is a public policy problem. Both are very challenging.

SPEAKER 1: Question from the audience?

ELLE WINEMAN: Yeah, Elle Wineman. On the interdependent privacy research, you assigned certain monetary values that people placed on giving up their friends' information. How did you derive those values? Were people asked what kind of value you place on it, or was there some other scientific method to come up with that?

YU PU: That's a really good question. So we used the methodology of conjoined analysis, so which is a methodology to evaluate people-- how people make trade-offs. So basically it assumes the app that is composed with different features that have different levels than by combining different features and different levels.

We can generate a list of apps that represents which have different informations the app collects about users. And we then ask the users to rank a list of apps. And based on their preferences, we

run a statistical analysis to infer or how much utilities they put on each levels of an app. And therefore we calculate the monetary value of these privacy.

JENS GROSSKLAGS: So one tiny feature is that one of the attributes is actually money. And having the relative importance of the different dimensions of what an app could be composed of can then be translated with the help of this one monetary dimension in the monetary domain for all these attributes. So in absence of having some price as one of the dimension, you could now directly translate it into the monetary dimension. But since we have price as part of the attributes, we can translate also then personal privacy, interdependent privacy, and various other kind of factors that we checked for also into the monetary domain.

ELLE WINEMAN: Thank you.

SPEAKER 1: Another audience question.

MARK WEINSTEIN: Thank you. My name is Mark Weinstein. I'm the founder of MeWe, which is a privacy-centric social network that gives people an alternative to Facebook. Tim Berners-Lee is with us. The founder of the web, he's our adviser.

I'm concerned when I hear this panel talk about all this automation. You're now going to automate my privacy decisions, based on my preferences. I think Yang, I think you spoke eloquently about the dangers of that. Shouldn't we be presenting people with factual data about what's happening for every app for everything they're using with their privacy, and then let them select.

The earlier panel, they talked about we need noise, we need critical thinking. We just saw an election where fake news-- we know that everything at Facebook is algorithmically filtered for what our preferences are supposed to be. And we know that our opinions change based on information, based on factual information.

So I'm wondering whether the panel thinks shouldn't we be giving better information rather than trying to automate the preferences, the privacy preferences, and the cookie preferences of somebody? Shouldn't we actually be giving them better information? Regulating that, and then having them make a selection really every time they join an app?

MAHMOOD SHARIF: I think that is correct to some extent. We want to give users to make their own decisions. At the same time, it might be infeasible.

Because a lot of decisions they might need to make are just too many. And they will spend their time just making decisions whether they want to be tracked or not. So we should try to understand what are their real and actual preferences, and from there, try to generalize.

MARK WEINSTEIN: Yeah, but you're saying that we should try and ascertain those. But again, this is actually the problem of privacy. This is the world's problem today. We're going to figure this out.

We're the technologists and I'm a technologist. But it's not my job to figure out what's important to my member. It's my job just to give them the information and let them make the decision.

That's also how democracy has always worked. So I think there's a very slippery slope that you guys are talking about. That we're supposed to somehow infer by their, like you said, past behavior.

YANG WANG: If I may add, I think that there is definitely value for some of these automatic decision making when the context where scenario are not particularly sensitive. Because again as he has said, it'd be too much as a user to make every single permission request decisions. That's just too much. That's not their primary [INAUDIBLE].

MARK WEINSTEIN: We should probably standardize. Maybe I don't think you standardize the way of filtering that decision matrix.

YANG WANG: So I think the research on the contextural factors are very important, because these could help us potentially identify what kind of scenarios where context would potentially be sensitive to the users. And then you can adopt some sort of tiered system where for these kind of low sensitive scenarios you could potentially just go ahead and make these automatic decisions. But of course, you can show user, hey, we made this decision for you. But for the more sensitive ones perhaps a focus should be more about providing information. That's my personal view.

MARK WEINSTEIN: As chairwoman Ramirez said earlier, everything that's happening is probably like your permanent record. This all becomes part of your permanent record. And so for us to decide what's of low importance when it all goes into your permanent record, I just think it's a slippery slope.

And I really admire the FTC for everything that they are doing to protect us. I think it's rather remarkable. We hope it continues.

SPEAKER 1: Thank you. So I wanted to touch on another point, which is the methodology that were used for these studies. There's a lot of interesting methods here, and we've been hearing more and more on this panel and others about crowdsourcing methods.

And you guys didn't have time to touch on it. But in your paper you had a really interesting approach to making sure your crowd workers were not slacking off. I wonder if you could just tell us a little bit about that.

YU PU: Things we've discussed that we ask people to rank a list of apps. And app differs in four dimensions. So we believe it would poses great challenges for participants to rank it. As a result, we find that quality is not that good. So what we do is that based on the results, we want to investigate whether our users have the rankings or the rankings has demonstrated that normal user behavior.

For example, we investigate whether the results indicated that people want a preferred to have an app that is free-- that costs maybe $2 then choose a app that's free while the rest of others of the app are the same. So if the participants indicated that they would choose the costly apps, in that case we would think that he did not do very careful. He is not very careful when filling out our surveys. And hence we will maybe fail to their data.

JENS GROSSKLAGS: I think more generally one really has to use a portfolio of approaches to encourage participation to motivate people to contribute to the studies in a meaningful way, to filter out at the same time those that do not participate meaningfully, or may even be using automated tools to participate in studies. And so for that, we really need a portfolio. So you indicated one of them, which is essentially an economic filter. So if people make economically very unreasonable choices, then we could assume that they did not pay a reasonable attention, because people would prefer not to pay for something rather than to pay for something.

We also introduced a pretest that helped us to identify individuals who generally seem to be more willing to participate in tasks. We also rethought some methods of soliciting preferences with different kind of practices to conduct conjoined studies that have been proposed in the literature. And have also had good success with that. So in general, I think the take away is when really should encourage replication of studies, and to try different kinds of approaches to understand whether what has been done is robust, whether the findings make sense, and always pushing the boundary of what can actually be done. And I think that's really, really important in the context of privacy where so many questions stand in the space where the experimental research really captures [INAUDIBLE] consumer behaviors and practice.

SPEAKER 1: Anybody else want to comment on some lessons you learned about methodology during your study?

MAHMOOD SHARIF: So I think that in our case, one of the most interesting parts was asking users about their own browsing history, which enabled them to reflect really deeply. For example, one of the participants was able to tell us how she's a fan of anime. And she was embarrassed if someone would see ads or targeted ads on anime shows showing up on her screen.

YANG WANG: Yeah, I think in our case, the drawing task was very, very eye opening. Because usually we ask people to describe what they're thinking. But once you kind of force them to draw that on a piece of paper, it really kind of nudging them to externalize their internal thinking. And I think that was very helpful, in our particular study.

CHANDA PHALEN: With our interviews, we had two phases set up. And the first was disguised as a user experience interview. And we never actually used the word privacy or primed them to think about privacy. So we just let them talk about it they cared, and we didn't talk about it they didn't care. And then in the second phase, we did tell them we were interested about their privacy and their privacy concern.

And the way that the answers switched when they didn't know that we were talking about privacy to when they did, all of sudden they started saying all the answers that they just learned

in their undergrad class about privacy. They were saying what they thought they were supposed to say rather than what they were actually interested in. And so the awareness of the subject and how much you're focusing on privacy is an important thing to think about when you're eliciting people's concern.

SPEAKER 1: So I think we're out of time. And next is lunch. Lunch is available for purchase in the cafeteria. You go out to the left and around. We're also having a poster session, which is in the room kind of right across the hallway.

I encourage you to pick up some lunch in the cafeteria and bring it to the poster session room. We have some tables and chairs set up there. I think we have some cookies in there. And we have some great posters and great researchers who would love to talk with you about their research. And we will all be back here at 2:30. Thank you.