FTC PrivacyCon
January 14, 2016
Segment 2
Transcript

KRISTEN ANDERSON: OK everyone, please take your seats. We're about to get started with the next session.

Good morning. I'm Kristen Anderson, and I'm an attorney in the division of Privacy and Identity Protection within the FTC's Bureau of Consumer Protection. I'm here to introduce our second session of the day, which is on consumers' privacy expectations. We'll hear from six researchers in four 15-minute presentations, and then we'll conclude with about 20 minutes of discussion where we'll identify common themes and ask the presenters about their work and its implications.

Without further ado, I'll introduce our first presenter. We have Serge Egelman of the International Computer Science Institute at the University of California at Berkeley. Serge will start us off with his presentation on Android permissions. Serge?

SERGE EGELMAN: Thank you for that introduction. So this is work that I've been doing with several students recently, where we've been looking at privacy and how private information is regulated on mobile platforms.

So to give you, I guess, a brief overview, most of this work is on Android, and that's only because Android actually has a pretty intricate permission system to try and implement notice and choice. So whenever an application requests access to certain sensitive data, it's regulated by this permission system. And so when users install an application, they see a screen that informs them of all the possible types of sensitive data that that application might be requesting in the future.

And so the question was, does this actually implement effective notice and choice? So do users understand these messages about how applications could be using their data in the future? So we started this project a couple years ago by doing an online survey. We had over 300 Android users, and we just showed them screenshots of these permission screens and simply asked them if an application was granted these abilities, what might that allow the application to do? We then followed that up with a qualitative study where we had 24 people come to our laboratory, and we interviewed them about similar concepts.

And what we concluded from this was that many people were simply habituated. Since these appear every time people install applications, not only does it list what abilities and types of sensitive data that application is requesting in the future, but all the possible types that it could request, even if the application never takes advantage of that. And so people become habituated. They see lots of these requests that have lots of different data types, some of which they don't understand, and therefore they learn to ignore these because there's just so much information there.

Another problem was that people were simply unaware. Since this occurs whenever you install an application, a lot of people said that, oh, this is just part of the license agreement, and we know that we need to click through that in order to continue installing the application. So maybe this occurs at the wrong time in the process. And since it happens after the user clicks install, it could be that they're already committed to installing the application. There are various cognitive biases that relate to this. And so therefore, it's unlikely that they're actually comparison shopping based on privacy, even if they wanted to.

Another issue is that understanding whether a particular application is going to access a particular type of data really requires a good understanding of this whole permission system and what are the different types of data that are regulated by the permission systems. So understanding whether an application is requesting a data type requires understanding the whole universe of data types that are governed here.

And so we made these recommendations, and what we concluded was that a lot of this could be taken away. So transparency is great. Notice and choice is good. But the problem is, when people are overwhelmed by the notice, which is what we see with privacy policies on websites, they eventually just ignore it all because there's so much information.

So what we found was that a majority of these permissions could probably just be granted automatically without showing the user lots of information, because either the dangers are very low-risk-- for instance, changing the time or causing the device to vibrate-- or they're simply reversible. So if an application does abuse one of these abilities, chances are the user can find out about it and simply undo it, and there's no lasting harm then.

At the same time, there are a few very sensitive things which, because of doing this install time that's probably the wrong time during the process-- the user has no context about how the data might be used in the future-- these could probably be replaced with runtime dialogs. But another open ques Tw 26.05 0 T(in)2(g)12rt9m ju disti that a-2(how)2(e4(qSe)-4(b-2(g)10(i)42(obe)-2(e)4(-x)-10(it)-2(a)4(

is an example of one of those indicators. It appears in the top status bar. And most people assume that the only time that GPS information is collected, this icon will appear.

And it turns out that's not true at all, and in fact the icon only appears in 0.04% of the cases where location data was accessed. And that's because every time an application requests location data, the operating system caches that for performance reasons, and also to preserve battery life. But then when another application accesses just the cached location data as opposed to querying the GPS hardware directly, this icon never appears. Similarly, applications can infer location based on cellular network data, nearby Wi-Fi hotspots. And it turns out, most applications are using those methods to infer location rather than the GPS hardware. And therefore, most of the time when location data is collected, people have no indication that that's occuring.

So having the notice and choice at the beginning when users install the application obviously doesn't work. We've tested that. The ask on first use that's currently happening isn't really working because of the different contexts in which users might be interacting with applications. So maybe we could have runtime requests all the time.

So every time applications request data, we can have a little notice appear. Well obviously, that's really impractical too. So the 27 million data points that we collected, that result in per person about 200 pop-ups per hour, most of which is due to requests for location data. But you can see that if there are other data types that were pretty frequently requested. And so having lots of pop-ups appear on the phone is not really a good way of going forward either, because that's also going to lead to habituation.

But at the same time, in our exit survey, what we found was that the vast majority of participants said that given the opportunity, they would have denied at least one of these requests. And on average, they would have denied a third of the requests.

So how do we do this? How do we give users control over the things that they actually care

data in the foreground and then ask the first time the application requests the data in the

purposes. So as regards to deletion, users predominantly expected websites to allow deletion of their collected data. But websites generally do not allow that.

So there can be other types of mismatches, as well. One example is a website specific mismatch. For example, users do not expect banking websites to collect health information. And most of the banking websites we looked at do not do so. However, there can be specific websites-- for example, Bank of America, which was one of the websites we looked at-- that indeed collect health information. So we can see, this is a mismatch that is specific to a certain website.

So based on the results of our study, we could come up with notices that have less amount of information than a full notice. For example, we looked at 17 data practices. A notice could show information about all 17 data practices, or we could show information about data practices where there's a mismatch between what users expect and what websites do, or actual data practices of websites. So for example here, for the Bank of America privacy notice, there were mismatches for 11 data practices out of the 17. So if you show only 11, that would be about 35% reduction in the amount of information that the user has to read and process.

We could also just show information about mismatches that are more privacy-invasive from a user standpoint. For example, I talked about the yes/no mismatch versus the no-yes. If you find that the yes/no mismatch is more invasive, we could only show information about those mismatches.

And in the case of Bank of America, it's only five data practices for which there's a yes/no mismatch. So that would be 70% percent reduction in the amount of information shown in the notice. However, the caveat here is that we do have to go ahead and test with users how effective the shorter notices would be. Yeah.

So as part of future work, we are planning to also study expectations in the desired sense and compare that with expectations in the likelihood sense, and also compare both of them to actually data practices of websites. We will also, as I mentioned, test effectiveness of notices that highlight mismatched expectations and see whether they actually reduce user burden and whether users can make better privacy decisions.

Yeah, that was all. Thank you.

[APPLAUSE]

KRISTEN ANDERSON: Thank you, Ashwini. Next we'll here from co-presenters Heather Shoenberger of the University of Oregon and Jasmine McNealy of the University of Florida. They'll be presenting on reasonable

importantly are our main dependent variables. So we had the always click Yes. So again, we're assessing behavior, whether or not the participant always chose to click Yes related to privacy policies or terms and conditions online.

And the second one was privacy concern measured on a three-item scale about whether they thought that data companies would collect information about them that would make them feel uncomfortable. And Heather's going to come and talk about some of the relationships we found.

HEATHER SHOENBERGER: Right. So we diverge a little bit here, where we're very positive about our findings. And also, I wanted to note-- well, I'll note that in a second.

So our always clicking Yes variable was our indication of behavior as our DV. This was an hierarchical regression, and I made it very simplified for this because we are under a time limit. The first block was demographics. The only demographic in this particular equation that was significant was age. And it's no surprise that it's younger people that predicted always clicking Yes. We've seen this in numerous reports, where younger people tend to be a little bit more careless online, maybe a little more apathetic, et cetera.

Then we move to the second block, and these were two variables that did come up in our survey that have also been used in numerous studies before ours. And social trust in this particular case was not a predictor, but control efficacy was. So even though they may not actually be able to control their data, the belief that they can predicted always clicking Yes. And we believe this is the result of the confidence that people have if they believe they have control, and as a result they go ahead and say, sure enough. I'm just go ahead and click Yes, but I'm confident and I trust that this is going to work out for me.

Those who had a negative-- oh, so the next block were all items that were derived from our interviews. Of course, some of them you've seen in previous studies as well, but all of them were derived from our interviews. So negative experience-- those who had had fewer negative experiences, self-explanatory, were more likely to click Yes without reading any terms of agreement, no further investigation.

Peer recommendations-- we were really hopeful that a peer recommendation would kind of be an if-then rule. If a peer recommends Snapchat to me, I will then go ahead and download it. That was not the case in our regression analysis. It wasn't significant.

Convenience was a pretty big variable made up of items like that policies are too long. It's faster to just skip them. They're full of legalese. Some of the information that we heard last night at the conference about how these policies are just laden with too much material for consumers to ingest, especially in an oversaturated environment with jobs and time constraints, et cetera.

And then the two variables that are really important to us for this study were both essentially cues. One was site appearance. If the site appeared to be safe and not weird-- it didn't raise any skepticism-- again, we've seen this in previous studies but our participants noted this in the interviews, as well-- predicted clicking Yes, if the site looked safe and also was familiar.

And then just simple presence of a privacy policy or an icon like Trustee also predicted clicking Yes. So this was our behavior. And at the conclusion of this we thought, we're on the right track here. These cues are what is driving the motivators of actual behavior online, and we were really excited.

Then, we got even more excited for our privacy concern variable, a variable that has been heavily researched in this area. Many researchers have noted the-- and the panel before us noted that there is a disconnect between privacy concern and actual behavior. We may have a potential to bridge that with this research.

So the [INAUDIBLE] regression is in the exact same format. Higher ages and higher education-- again, no surprise-- predict privacy concerns. Lower social trust, the trust of the institutions, predicted privacy concerns, lower control efficacy, both in line with previous research. People who had suffered more negative experiences were more likely to say that they had higher privacy concern. Again, peer recommendation-- we'd had high hopes for that, but it didn't work out.

Here's the catch for people who are in the advertising industry who are in the business of collecting and using consumer data. They would have to adhere to those guidelines in order to use the cue on their sites, which would signify safety, increase trust, hopefully, et cetera.

So we would also do research on what icons would be most effective to consumers, and also link those icons to readable policies. Another thing that we noted was the convenience variable was made up of items like it's too long, it's full of legalese, we don't understand. And if we could m6.1ade upth2(d)6ld; TJ4d ab t t th2(dis004 Tw T* 3(m)-6(ak3Td [9.36/P <</M--ID 1-1.15 Td [0.67/P <</M

have to realize is genetic data is the most personal data there is out there. Not only is it a unique identifier of us individually, but because of the familiar nature of DNA it can also identify our family. So when we're talking about privacy in this context, we're talking about it in a much broader context-- not just personal, but looking at the family.

We also know that this data is inherently identifiable. OK? There's growing recognition that it is simply not possible to de-identify this data in a way that makes it impossible to re-identify it. It may take a good skill set, but as we get increasing numbers of genetic databases out there, as there are more public databases, we know that we can re-identify that data.

The other thing is, this data is irrevocable. If there's been a privacy breach, you can't change it. It's not like your iTunes password. You can't come up with another one. OK? So this is a different type of data. OK?

Does it matter if this happens in a direct-to-consumer genetic testing situation? Well, the first thing we have to realize is the difference between traditional genetic testing and what happens when we have genetic testing in a direct consumer setting. Traditionally, genetic testing has happened within a country's health care system.

And that's important because when an individual gets the genetic test in their health care system, they're deemed a patient. And by being called a patient, that enlivens a whole host of professional and regulatory oversight, existing legal duties of care, and simple things like doctor-patient confidentiality. So all the government systems for data protection of health care kick in, because that's a patient.

When we look at direct-to-consumer genetic testing, we have to realize that at its core this is a commercial transaction that occurs in each country's marketplace-- and increasingly, in market space, because the majority of the activity is actually online. When an individual engages with DTC, they engage as a consumer. What that means is that enlivens each country's consumer protection legislation. It also enlivens some particular legal protections in contract negligence, et cetera. OK? But a very, very different situation.

What does the general public think of when they think of privacy? At the Center for Law and Genetics at the University of Tasmania, we've been looking at genetic privacy issues for the last 20 years, and in the last few years we've moved into DTC. Some of our early research in direct-to-consumer genetic testing suggested, from the Australian general public's perspective, that privacy concerns were going to be the key constraint on commercial uptake. Interestingly, this past year we found the same results when it comes to intention to biobank-- in other words, giving a genetic sample into a genetic database for nonprofit, institutional, and health-related research as opposed to commercial.

We've also modeled the DTC space. And that was an interesting exercise and forced the thinking to go broader than just the consumer-company interaction. What we realized very quickly was not only does DNA go a lot of places-- that sample travels from labs to companies and who knows where, through the postal system usually-- but also, those results can go places. OK? The actual genetic data about those individuals gets spread around.

ANDELKA PHILLIPS: Well, I've actually been looking at the contracts and privacy policies of direct-to-consumer tests or companies that offer tests for health purposes. Now, as has been noted in the previous session and also in the previous group's work, these contracts and privacy policies appear everywhere online. Basically, any website you use, any software update you make, will be subject to terms and conditions. And they'll be presented either as terms and conditions, terms of use, terms of service, privacy statements, privacy policies, and sometimes in this context they're combined in one document. At present, these are used to gather not just the purchase of DNA tests, but also using the website and sometimes participation in any research the company is doing.

Now, as several people have previously noted, people don't tend to read these contracts and privacy policies, partly because there's just so many and it would take too long. This industry is

And I noticed three common themes in your answer to your findings. The first is that notice

As we walk into this, there's definitely a lot of discussions over different expectations versus privacy, or people not understanding the legalese in direct-to-consumer genetic contracts. But is that a public policy problem? I'm not so sure. Let me draw an analogy. Say I'm not necessarily sure what goes into my Chipotle burrito. Sure, I'm able to pick different fillings--

SERGE EGELMAN: E. coli?

ALAN MCQUINN: --I may be able to pick different fillings, but I'm not so sure how they're sourced. So when you ask me questions about what's in my Chipotle burrito, my expectations may differ from the reality of what's in there.

Now, that's not necessarily a public policy problem, right? But what is a public policy problem is when consumers start to get sick or have food posing as a result of the contaminated food from a Chipotle burrito. But when I'm listening to these presentations and reading these reports, I'm saying that we're talking about what's in the privacy burrito rather than actually talking about privacy food poisoning. That's just some food for thought, I guess, and I look forward to a good discussion. Thank you.

DARREN STEVENSON: I have no way to connect to the burrito, but we wish Chipotle well with their current issues. I do.

[LAUGHTER]

So at the risk of stating the obvious, I think what we have here is we have evidence, empirical studies, that show that consumers have expectations. All of you in this room, you guys are not ordinary consumers because you're here at Privacy Con. But ordinary consumers, we're seeing that there are consistent, measurable expectations. I really enjoyed the studies, and I encourage you all to read them if you have not read the papers.

So how can policy, which tends to move slowly, track and be responsive to something that is changing, that is dynamic? So if we were to have Privacy Con in three years, next year, five years, and we repeat all these studies of consumers, would we see the same expectations? So how can policymakers incorporate this sort of moving target of consumers' expectations?

So I look forward to our discussion here, and I think we can open up to questions?

KRISTEN ANDERSON: Yeah. Go ahead.

DARREN STEVENSON: Or if you have any responses to our comments.

JASMINE MCNEALY: I don't know. I like the burrito analogy. But at the same time, if Chipotle has lean steak or whatever they have, I mean, if they make representations to the consumer that it's from a certain source, then you have expectations that, hey, my beef is from a certain source. And even if we don't know exactly where it's from, we have an expectation that we should get at least a product of some, I guess, quality. Or at least, we expect that regulators would enforce the restaurant giving us a product that either won't make us sick or won't have had something done to it by a worker there, right?

So I think there is a certain level of protection we expect from re wrrsit respect t4(t)-2(hi)-2(ng)11(s)-1(( )]TJ T

S8-6(o)-44(r)-10(g)6(u)-4(l).1(at)-6(o)-14( )]TJ 0 Tc 0 T410.91 0 T[sp2(a)4(r)4(t)-2(i)-1(c)4 wryp,4onsty ctl2æ
An(w)2(e)4( )]TJ T* 0(e)4(x)-10(pe)4(c)4(t)-2k thecarhesk tve w atverst al2etcpee w whoul1b(ke)4j(t)-a(t)-12

ghi itsrctiheteares tp3(r)3tomethee fromhetsales avehee atitesS( )]TJ2-2491 -1.15 Td 0(I)23(t)-2(hi)-2(,n(f)3(r
S8(E(h)-Ror)-G(t)-E(h)- E(h)-G(t)-E(hI)2LttS8(6(o)-60(922no)-1g)6-uo s5(0(s410-2(at).1-6(h)-e i(e)-0(s410-0(s
Sf(r)-34(r)0(hi)-21(s)-4(t)-a(e)4ch4cS8((he)0d4(i)-d1ahe)04(s)-1(t)-ud2(y)22nolvinghat sigeo

And so we wanted to see whether making that more apparent to users-- so trying to highlight what types of data might be collected by those websites from your Facebook profile-- we expected that that would have an effect on whether people use this. And we found that was not the case. And when interviewing subjects, they said, oh, well, they just assumed that Facebook is giving away all this data anyway, so I might as well get a benefit from it.

And so that's sort of the learned helplessness issue. I'm not sure there's-- I think addressing that part of it is sort of putting the cart before the horse, because I think one of the issues we need to focus on are the expectations before they're formed. Some of that might be doing a better job of public education with regard to online privacy. Other pieces might come in the form of enforcement, making that somewhat more subjective.

So yes, the law moves very slowly. Technology moves quickly. But I don't think the issue is making the policies around specific technologies. The issue here is narrowing, or closing, the information asymmetries. So while we don't expect people to read every privacy policy that they encounter, we have some expectations about what a business might be doing, as was pointed out. So regardless of what they say about what farm the beef came from, I don't expect it to have E. coli in it. And that's not something that they need to explicitly provide notice for. It just should be expected that there's no E. coli in this beef. I'll leave it that.

[LAUGHTER]

ANDELKA PHILLIPS: I'd like to say-- because we kind of ran out of time a little bit-- there is really a need for more transparency in the industry we're looking at, because often if you look at website claims there will be quite a gap between what the contract actually says and what the website is encouraging consumers to believe when they are encouraging people to purchase tests.

And the other thing is that, because the industry is so new and the technology is changing so fast and it is largely unregulated, a lot of tests that are coming to market haven't been validated. So there is a question sometimes about what the consumer is actually buying, because the values to the company is the sequenced DNA, which they are using in ongoing research often. So they're selling a product that gives them very personal data that they use for a long time and may not be destroying ever, potentially. And the consumer, an ordinary consumer, doesn't necessarily have the expertise to understand all of the risk.

And the other thing is that genetic test results are complex in nature. A lot of general practitioners have trouble interpreting genetic test results. And there's been some studies that have shown that a lot of GPs wouldn't be comfortable with interpreting a DTC test result if a consumer brings it in. But at the moment, most of the time it's being avowed as a consumer service.

And in terms of particular worrying terms in contracts, in some countries, like the UK, the Office of Fair Trading, which has now been disbanded but is the Competition and Markets Authority, has a history working with industry to try to discontinue certain unfair terms, as well. And that's what I would say. There are some terms that really shouldn't be in the contract, because it's

making it a very unfair and unbalanced bargain. And a lot of the use of these contracts is also eroding traditional contract law principles, really.

And I think people will often tend to engage with these-- and I think your work shows that-- much more differently than they would with a paper contract. So if a browse [INAUDIBLE], which is where the terms are on a hyperlink, it's akin to walking into a shop and being bound by a sign on the wall that you didn't see and walking out again. And that's really problematic. Thank you.

DARREN STEVENSON: Yeah, I think I'll add on. So on someone's slide there was a mention of incorrect mental models. And a lot of us think through consumer knowledge, and I think the educated consumers. So no one would argue for an uninformed consumer as the goal, but I think I want to push back a little bit on that idea, that our goal or the goal of some of this work is to correct mental models.

I'm curious what you guys think. So someone smarter than me said something like, all models are wrong. Some are useful. And I think that consumers sometimes have very wrong or inaccurate models that are helpful heuristics. I'm curious if you guys in this work, since you're all studying consumers' perceptions, see those inaccuracies actually beneficial or not that we want this inaccurate model? Does that make sense?

SERGE EGELMAN: Yeah. I think that was my slide. I think one of the bigger problems with notice and choice is that there, I guess, is unreasonableness on both sides. So there's unreasonable expectations on what the consumer should know to make an adequate choice based on the notice given to them. So it's unreasonable to expect every consumer to read every privacy policy that they encounter.

At the same time, yes, people have really bad mental models about what's happening with their data when they go online. And I think maybe there needs to be some better outreach on that issue. But at the same time, I think then that goes sort of to enforcement, which is instead of thinking, well, did the company give notice, and was it incorrect and outright misleading? But also, adding into that equation, is it reasonable to expect that someone could actually understand this? And I don't think that's currently being taken into account.

HEATHER SHOENBERGER: I'll also answer that very briefly. As far as using heuristics are part of the cognitive miser sort of mental model, first of all, I disagree that heuristics and using them are faulty. They're almost always correct. I mean, we rely on them all day long in various capacities. I think what we were arguing for were heuristics that were actually backed by informed and concise and true information that the FTC approves.

And so by promoting consumers and allowing them to see what these heuristics mean, promoting the cues to those consumers, it gives them a meaningful choice. So the heuristic is no longer something that risk is as much of an issue for, more as something that they can genuinely rely on as an indicator of safety.

KRISTEN ANDERSON: Heather, can you guys talk a little bit more about how you see that kind of a heuristic coming into place, how you would develop it based upon an average consumer's expectation? Given that we heard a lot of the findings are consumer-dependent-- it kind of depends on your background, the experiences you've had, your age-- how would you go about trying to develop something that would be generally applicable?

HEATHER SHOENBERGER: We are in the preliminary stages of doing that. And this would be something that we would be testing in a lab probably, a physiological lab, looking at people's automatic responses, in addition to self-report. But that said, looking at heuristics and making cues that were in line with guidelines that we have come up is based on consumer expectations of the different types of data collection.

So we came at it-- and this is an arguable point-- from the type of data collection and how it's being used, and then entities could opt in depending on how they were collecting and using that particular type of data. So there would almost be a continuum of badges or icons or cues that you could use. And then in order to use that on your site or within your materials, you have to adopt the FTC's guidelines that went with that particular icon. And we would empirically test every single element of that.

So the icon itself might be something that we would have to test to see if it was something that caught someone's eye. Someone noted-- I think it was [INAUDIBLE] that people didn't notice some of the privacy policies. That's something that we could correct with better web design and better icon design.

KRISTEN ANDERSON: We have about 20 seconds left, so I'd like to give you guys an opportunity to ask a last question.

ALAN MCQUINN: So to follow up on what Darren said with how privacy concerns kind of have morp

drug-- this is now being monetized, and this is now in the corporate sphere. And our protections