FTC PrivacyCon 2017
January 12, 2017
Segment 4
Transcript

KRISTIN KRAUSE COHEN: You came back from lunch and that there's a lot of you here. We have a great panel ahead of us on online behavioral advertising. And I hope many of you were able to check out the poster session. And thank you to all of those folks who put up posters. I found it really informative and really appreciated that.

This panel is going to be run like the other ones, where each of the presenters is going to tell us about their work. And then we're going to have a period of moderated discussion. And so we're going to go ahead and get started.

James Cooper is going to be our first presenter. He is a law professor at George Mason Law School and he previously worked at the FTC as the Deputy and Acting Director in the FTC's office of Policy Planning, as well as serving as an attorney adviser for Commissioner William Kovacic. So please join me in welcoming James Cooper.

JAMES COOPER: Perfect, thanks, thanks. Thanks, Kristin and thanks to the FTC for inviting me to present my work here. I think that this is, PrivacyCon's such a fantastic thing. I just think this is a, I can't say enough good things about the idea of bringing in all this academic work and having it melt with policymakers here at the FTC.

Because this is what the FTC is about, I think. So anyway, I'm going to talk today about anonymity, autonomy and the Google privacy policy change 2012. So privacy is a multifaceted thing. We could spend all day talking about it. It means lots of different things to different people.

But at some level it's about controlling the information about, information flows about yourself to the outside world. One way to accomplish that is to retreat from the world. You can seclude yourself and, but the other way, you may want to interact, you may want to interact with the world.

You may want to engage in the world. But without your actions or your words, whatever you're doing being actually traced to you. So you basically put a mask on. Put a mask on like one of these guys, OK?

So the guy on the left, we know who he is, right? He doesn't have a mask. We know that Johnny Utah, football star, an FBI agent. On the right, who knows who that guy is, right? He's wearing a mask. I don't, are you guys too young? So he's is wearing a, if I would have slides from the remake, maybe this would go over better.

But the guy on the right, he he's wearing a mask, OK? So he can engage in activity in the outside world. He's engaging, but we don't know who he is, right? I won't let anyone know who he is. I won't give it away.

Now he's wearing his mask so he can rob banks to fuel his adrenaline-powered lifestyle. But others wear a mask in anonymity to, we engage in anonymity. One of the things we get from that is autonomy. There's a zone where we can engage in behavior, we can make decisions free from observation, free from interference.

Autonomy, this notion of privacy, this zone of autonomy underlies a lot of 14th Amendment jurisprudence. It goes to reproductive rights. We're given a space where government can't intrude on personal and intimate decisions. So autonomy is very basic in our constitution.

And there are also some utilitarian reasons we may want to engage, to protect autonomy. You know, first of all, you may, if you're being watched all the time, if you can't be anonymous, you just may be embarrassed about what you're doing. So there is dignity harms.

If you're being watched and this is an important thing and what the paper goes to, you then may refrain from engaging in certain things that are crucial to your development. Things, everyone needs a zone to explore or to figure out who they are. So if you can't engage in that personal development, that's going to retard your growth as a human.

That can also have the spillover effects on society, right? I mean, so this idea of self-discovery has a lot to do with creating variety and which some argue may be crucial to a functioning democracy. So we need this kind of variety.

And finally, if you can't be anonymous. If you can't go check that book out at the library and with the mask on, maybe you won't do it. Maybe if you've got some condition that you want to learn about, but you're embarrassed that someone will know, you'll engage in what's called privacy-protective behavior. That also is harmful.

So there are a lot of reasons why, this concept of anonymity, which is one of the concepts of privacy, is important. So what am I doing here? What is the study about?

Well, we've seen from other panels and people who are familiar with the privacy scholarship, one of the big questions here is to figure out how consumers make trade-offs with privacy and other values.

Now, there are a lot of work out there that tries to get at that directly by conducting experiments to figure out how much people are willing to pay or are willing to trade off in some sort of hypothetical trade between money and privacy.

Well, what I wanted to do is to go and look in the real world and it's really hard. And one reason a lot of this work is done experiments is it's kind of hard to find natural experiments or find real world experiments where privacy's been changed and gauge consumer reaction. And that's what I wanted to do.

I wanted to, even though I'm not measuring directly the valuation of privacy, I wanted to measure how people change their behavior because of a reduction in anonymity. And specifically, I'm going to look at the 2012 Google privacy policy change.

In March 2012 when early 2012 Google announced that in March, on March 1, they were going to start combining data across platforms and have one privacy policy. And so, at the margin we think this reduction in anonymity is going to deter engagement in search behavior that you otherwise would want to keep private.

So some of my anonymity's going to be stripped away. So I may not engage in behavior that I would have otherwise if things would be private. So what I'm doing is I'm hopefully going to try to indirectly measure a reduction in autonomy due to this, reduction in anonymity.

Essentially, I'm looking at the manifest censored self. So I'm not searching for things that I otherwise would. And we'll go into a little detail of how we designed the study, but first kind of a threshold question is, did people know.

If I'm going to use the Google privacy policy change as the catalyst of sort of my experiment, the treatment in this experiment it's got to, people have got to know and care for it to really be relevant. Now I know that in this room probably lots of people knew about that or cared about that. But I needed more than that, just the privacy policy community.

So I went out and did some research and in a lot of this is in the paper but, surprisingly, I have to say. You know, I thought about this and thought well, if there is something in this little echo chamber that we have here. But no, there's actually all the major papers picked this up. Some multiple times.

As we know, the privacy advocacy community there was a lot of active opposition and there's a lot of you here who work at the FTC know, EPIC actually sued the FTC to get them to sue, to get the FTC to sue Google as a violation of the Google Buzz consent decree.

All right, so there's certainly, there is a lot out there. There is a lot in the press. This wasn't just like a little quiet, quiet thing that was on specialized, you know, tech blogs and things like that.

Also, here's another bit of evidence a little, a little more data here. This is Google Trends data, and I'll explain a little more about Google Trends in a second. But of a search for Google and privacy from 2011 through 2013, you see it's basically flat until right around the announcement.

So a lot of people, to the extent that Google Trends in is, and there's a lot of research to suggest that Google Trends does a really good job of capturing what people are interested in throughout the country. To the extent this does, you see a spike between January and March 2012 when this was going on. So there seems to be a lot of interest.

All right, so what's my research design? Well, basically I'm going to look at, I'm going to look at sensitive, what I call sensitive search relative to non-sensitive search in a difference-in-differences framework. And I'm going to use, the variable I'm looking at here is Google Trends data. I just put that up now.

Google Trends is something that Google constructed from 2012, I'm sorry 2010 through present day, that's an index. And it's publicly available. You go, Google Google Trends and you'll go to their page and you can see all this metric of search groups.

Now it's not a search volume, it's an index of relative popularity of search terms and somewhat complicated and somewhat opaque exactly how it's done. But even though it doesn't capture direct search volume, it does accurately capture trends.

So if a search term is going up or going down, you'll get the directions right. You may just not, the magnitudes aren't there because it's an index that ranges from 0 to 100, all right?

So I'm going to do this in a difference-in-differences framework. And what does that mean? It means I'm going to have a treatment group and a control group and I'm going to look at them before and after the experiment. The experiment here being the 2012 policy change.

So in this experiment, the treatment group is sensitive search. The kind of search that I may be deterred from, that I think that this reduction in privacy is going to impact. All right? And the control group is non-sensitive search. Things like weather. That I don't care if people know I'm searching about weather. I may care if they think I'm searching about porn.

So what you do in a difference-in-differences framework is you compare the control and treatment group before the treatment. And then you go back after the treatment. And if the hypothesis is right that the treatment, and here's just an example of some of the terms that we looked at.

You should see the kind of ads that my RAs get now. It's kind of true. Actually got a complaint about that. But, not a formal complaint, just, hey, I get some weird ads now.

Anyway, so the what I'm going to look at here is the difference between these two. And if mingling the data across platforms actually deterred sensitive search relative to the control group. The treatment group is affected.

This delta delta hat, which is my difference-in-differences estimator, should be less than zero, OK? So that's what I'm looking for. I'm looking for a nega-- the gap to increase. Which would mean this would be less than zero. The gap between sensitive and non-sensitive search got larger after the treatment, all right?

And so what do I see? Here's some results. I only have a couple more minutes, but the first, here's just some means. And what you see here is that that's the pretreatment group, the sensitive search is the blue bar, red bar. So you see a gap 53.5 and 58.7. So you got about a five point gap there.

After the treatment the gap actually shrinks, OK? So that is not, that doesn't suggest that the treatment had any impact. The gap shrunk, but of course, these are just averages. There's a lot going on, so I look at this in a regression framework, as well.

And basically the unit of observation is the Google Trend score for each of these searches. I have 20 sensitive and 20 non-sensitive searches. And for a week in a state, depending on the window I look at I have between 13,000 and 108,000 and I use lots of different controls. Week, term, state, fixed effects, as well as I search specific trends, things like that.

So what are the main findings? Well, I find in a short window, basically around March, and there are some things that I've done some work since that is not in the paper that was submitted. So it'll be coming out probably in the next couple of weeks on SSRN with some updated stuff.

But I find depending on the window you look at and the controls you put in, somewhere between a 5% to 10% reduction in sensitive search relative to non-sensitive. It's statistically significant. It's standard levels. In a short window, most of it comes from the month of March. So if you expand the window, you can get a larger, you can get an effect.

But it's almost all being driven by reductions in March. Larger time-frames you don't find any reduction. And one thing, if I had, that I felt was actually kind of interesting as I broke it down by states based on privacy demand, which comes out of the looking, creating an index based on privacy statutes in the state.

And I thought that, OK, well I'm finding this aggregate, but if I looked at these high privacy demanding states, say California, Delaware, Connecticut, places like that that have a lot of statutes on the book, I would find a big difference. And actually I didn't.

I was really surprised. I play with the data a lot. You'll see that in the paper. And there's really no difference in this. Here is just some robustness checks. This is the distribution of parameter estimates.

And what it means is I just randomly sampled out of 100 terms, reran the regression 100 times and looked at the parameter estimate and got a distribution. So this is the sampling estimate. And it lends credence to the point estimate I got.

And the main results that you see in the large window is centered around zero, which means sort of the average estimate there is zero. The short window is around negative two, which is about what I was getting.

So it suggests that the results are not sensitive just to the terms I happened to pick. That if you do a lot, again this is 100 random regressions and this is the estimates. Let's see, in the interest of time, I'll just go ahead and conclude.

So basically what I found is that I did find direct, find some indirect evidence here that there was some kind of reduction in the ability to be anonymous when you search. And that actually lead to reduction in sensitive search. Some people seem to have been deterred, at least in the short term, from engaging in sensitive search.

But again, this effect faded quickly. It's in the paper, but if you look at market shares and things like that of search engines, nothing changed. In fact, Google market share went up a tiny, tiny

amount during the time. So it's not like you see a lot of people fleeing to go there. It fades quickly.

Limitations on this and future work will, as I said, trends aren't volume, so all we're looking at is directional. We're not really looking at, we don't really have actual magnitudes. That would be interesting. Certainly I tried to do a good job with the search terms, but these are I don't have the universe of sensitive search terms.

We can find lots of sensitive search terms and there are probably other unmeasured margins that I'm not getting at that may be more important, or maybe you'll find a bigger impact. For instance, content in Gmail. Maybe I write less sensitive things in my mail. I'm not going to get access to that data.

Or viewing YouTube. You know, I'm watching less sensitive YouTube videos, right. I'm not going to get out get access to that. So those are some shortcomings and maybe some future work but, I apologize for going over, but thanks and I look forward to the Q&A.

KRISTIN KRAUSE COHEN: Thank you, James. And our next presenter is Steve Englehardt. He is a Ph.D. Student at Princeton and he actually presented last year on his open WPM tool. I think the chairwoman mentioned it this morning in her remarks. And he's going to tell us some of the measurements he's used, he's made with it.

STEVEN ENGLEHARDT: Thank you, Kristin. And happy to be back this year. So I'm Steven Englehardt. This is joint work with Dillon Reisman who is also in the audience and you may remember him from this morning. And Arvind Narayanan, both at Princeton University and the Center for Information Technology Policy.

So I want to share with you some insights from our most recent measurement of online tracking on the top one million sites. And as I'm sure many people here are familiar with, the web is very interconnected.

When you visit a website, when you load resources from that website, you're not just connecting to that server, but you end up connecting to a bunch of different third-party servers.

In particular here, when you visit CNN and the New York Times, you end up connecting to 84 different third parties. And any third party shared between those sites would have the ability to track you in some way. So we wanted to better understand this ecosystem. And we built a tool to do so. It's called OpenWPM.

You can check it out at the link here on the bottom of the slide on GitHub. And I presented on this last year, so I'm not going to focus too much on it this time, but I'll just say it's a web crawler that uses a real browser.

And we go and visit sites and collect anything we think might be useful for understanding what kind of tracking is happening on those sites. So we used that tool to run the Princeton web census, which is a monthly million-site crawl.

And there we collect things like JavaScript crawls, all the JavaScript files that are loaded on the site to look at things like fingerprinting, as well as request responses and browser storage like cookies to see how our fake users, or bots, are being tracked when we visit different sites.

And so the paper has a ton of awesome results. I encourage you to read it. Again, there's a link here at the bottom of the slide to check it out. I won't have time to go through all of the results. Here's a highlight of some of the ones I won't be touching on today.

In particular, showing that many, many third-parties are involved in cookie syncing. And also that the protection tools, like Ghostery or Adblock Plus do pretty well, but they end up missing things like less popular third parties or less popular tracking techniques. And we go into a lot of detail in the paper on that.

But instead I want to focus on a couple insights that I think are most relevant to the audience here today. And the first is the consolidation of top trackers. So if you go to the top million homepages and you've been the third-party requests by the organizations that own those domains, you end up with a graph that looks like this.

So as you can see, there's a heavy consolidation on this end of the graph. There's only a couple of companies that have a large presence and in particular, Google has a very large presence. Much larger than the other third parties. They're present on 85% of sites that we've visited.

And I think the takeaways here are that enforcement efforts can focus on these larger players and help set tracking norms. So if Facebook started to do fingerprinting, and just for those of you who may not be familiar, fingerprinting is a notion of identifying a user's device by the device properties.

Instead of say, setting a cookie on the device, and using that to track the user. So whereas you can clear cookies, you can't really clear your fingerprint because you can't change your device properties on demand.

So let's say Facebook started to use fingerprinting. You can imagine many other trackers in the long tail might feel it's OK to use it. So focusing enforcement on those top parties can help influence the rest.

Also that large trackers can quickly deploy techniques to a very large number of sites. We found a couple of new techniques during our measurement and we were surprised to see them on like 6% of sites.

But that's really because it was just one third party who decided to implement this tracking technique. And now all of a sudden it's on many sites, most likely most users would hit one of them.

And then also acquisitions can quickly shift tracking capability or say could combine data sets that may not have been combined before. And in particular here, Oracle would be an example of

this, right? They purchased BlueKai, they purchased AddThis, and now both of those tracking data sets you can imagine being combined.

The second point I want to make is that trackers can impede HTTPS adoption. So the HTTPS is basically a secure connection to a server. If you want to make sure you're connecting to that server and you want to make sure that the data you're getting back hasn't been tampered with, you'll do it securely through HTTPS.

And if that server happens to load resources from a non-secure server, it's called mixed content. So basically, some of the content on the page isn't secure.

And if that happens, some browsers like Firefox will show actual, will show a warning, a downgrade to the security indicator that in my opinion looks worse than just a normal HTTPS page which wouldn't have any kind of yellow warning on it.

So sites which may not normally want to adopt HTTPS because they're not handling credit card data, they're not handling log-in data, may actually avoid adopting it, so they don't get errors on their page if they happen to have mixed content. Or if, for example, a third party happens to include mixed content.

And of the sites that we found this mixed content on, half of them were caused by third parties, 10% by trackers. And it it's important to note that some of these parties actually did everything right. Some of the sites. They included only secure sub-resources and then one third-party cookie synced or basically redirected to a non-secure resource.

Meaning the whole page then had a security warning just because of the actions of one specific, one of those 84 third parties as I showed in the first slide. And we also found that half of all third parties are HTTPS only, so there's a lot of efforts needed to move the ecosystem to HTTPS everywhere.

And so the takeaway here is again, we can this, that tracking may have second order privacy impacts, and that may end up doing things like slowing the adoption of encryption or for non encrypted connections, actually leaking identifiers in the clear to the network. And for example, could aid in network surveillance efforts.

And the last thing I want to discuss is the use of new browser features for fingerprinting. And so what we found here was, all of these techniques that I show on the slide are being used, to fingerprint users. And in particular, Canvas is being used more often than these others, but all of them have some use.

And I think what I want to highlight here is any new browser feature that comes out, I expect it to be analyzed and potentially used for fingerprinting. If it will help someone identify that device, you can bet that someone will implement some technique to fingerprint users with it. And we can expect them to be used in that way.

And I want to focus on the fact though, that we're seeing browsers respond to this. So a year before we did our study there was a paper called The Leaking Battery, which actually looked at the use of the Battery API for fingerprinting. And then we went and measured the use of that in the wild and we found that sites were actually doing it, or scripts were actually doing it.

And in response to both of those papers, both Firefox and Safari ended up removing the Battery API from the browsers. Firefox unshipped it and Safari just never shipped it.

And I think the point here is that browsers are starting to view fingerprinting as abuse. They attempt to mitigate it during standards, during standardization, and now we're also seeing the removal of APIs that are being used for fingerprinting. And so that's kind of, I think from a privacy perspective, that is a positive thing in the ecosystem.

And we've also found that early detection of the use of these techniques can help stem adoption. So back when we first measured canvas fingerprinting it was being used on the top 5% of the top sites.

And then after we released that study, we saw it drop significantly, although in the absence of measurement we've seen it kind of coming back and there are more details in the paper on that.

And so our data is available, if you're interested in checking it out. The data used in this study, the million site measurement, as well as a bunch of kind of targeted measurements for different things like, let's go run a crawl with Adblock Plus enabled and see what tracking looks like.

All of that is available on the website, again at the bottom of the slide. And you'll see some links in later slides. But if you're someone who maybe isn't very technical, not used to writing SQL queries, getting some information out of that data might be pretty difficult. And this is an example of just the code that it takes to get third party responses from our data.

So you let's say you want to know what are the third parties on the New York Times. You'd have to write kind of the equivalent of this SQL query and data. But we don't want you to have to do that. Instead, we want you to be able to download some library. We're calling it Census.pi. That might change.

But we want you to be able to make queries like, give me all the third party requests for this domain. And then you can just pass in a pointer to our database, pass in the site you're interested in learning more about, and get back a bunch of rich results on that.

And we're hoping to do, we'll have a pretty thorough API, in particular, looking at cookie syncing, tracking and so on. But we encourage you to reach out to us. We kind of have like an alpha analysis service server set up. We're interested in giving people access to it. We're interested in seeing what people want to get from our data and playing around with the library.

So we encourage you to reach out to us. You can see my contact info here to send me an email and we can give you access to that data. So thank you very much. Looking forward to the panel.

KRISTIN KRAUSE COHEN: Thank you so much, Steve. I'd like to introduce Zubair Shafiq, who is from the University of Iowa, who's going to tell us about anti ad blocker.

ZUBAIR SHAFIQ: Thanks, Kristin. So my name is Zubair Shafiq, and this is joint work with Zhyiun Qian at UC Riverside, who is also in the audience. Today I'm going to talk about our research about the arms race between ad blockers and anti-ad blockers.

So before I start I wanted to give you some background about ad blockers and why ad blockers have become very popular recently. So online advertising plays a critical role in allowing the free web. But it return for this free web, there is a quid pro quo assumption that users agree to watch targeted ads to support these free services.

The annual online advertising revenue in 2015 in the United States alone was more than $60 billion. Unfortunately, this economic magnetism of online advertising has made it an attractive target for different types of abuses.

So for example, many of these ads are very flashy. They interfere with organic page content, which annoys users. Secondly, there is a security concern here, as well, because many hackers are using advertisements to actually send malware to a large number of users.

And these difficulties are in large part due to the complexity of the online advertising ecosystem. Where publishers have very little control over which advertisers will serve ads on their websites. Unfortunately, this complexity is driven by the need to serve targeted ads.

To show targeted ads to users as we just heard in the previous presentation, advertisers track users across the web using cookies, beacons and various fingerprinting techniques, which are not very visible to end users.

The online advertising ecosystem not only lacks transparency for end users, but unfortunately, does not provide any meaningful control to users to limit this tracking and the use of personal information. The users want control. Users want more privacy.

According to a recent Pew Research survey, more than 90% of users strongly agree that they want to know what information is collected about them and who is collecting this information. But again, unfortunately, more than 90% of users think that they have actually lost control. So there is a feeling of giving up.

So users do care deeply about their personal information and the regulators and companies know about this. There have been some attempts recently to do some regulation in this space.

I think roughly around 2009 the FTC started this push for advertising industry to do some kind of self-regulation. And I think this was a golden opportunity for advertisers to avoid heavy handed regulation. Unfortunately, this self-regulation has not proved very successful.

So for example, there was this ad choices program, which was started in 2011. This program allows users to see what kind of ads they are shown and why they are seeing a particular ad. But

unfortunately, this program does not give users any explicit control to opt out of the tracking which enables these targeted ads.

Moreover, many publishers and advertisers are not part of this ad choices program. There was another interesting effort called Do Not Track, which was in 2009.

This was supposed to be another voluntary program. It was supported by major browsers, but unfortunately the online advertising industry did not catch up and they did not support this voluntary opt out service.

Very recently the FCC has passed some regulations barring broadband providers to monetize user information without explicit opt in. And this is, obviously, very commendable.

FTC has been, I've heard, interested in something similar, but the recent political climate, uncertain political climate makes it unlikely that any media regulation would be passed in this regard.

So really the privacy conscious users and the research community have started to look at some technical countermeasures to get, to solve some of these privacy problems. There are these privacy enhancing, privacy preserving tools, which are commonly used.

For example, Privacy Badger, which specifically targets publishers that do not respect, do not track. Or Ghostery, which is a proprietary tool to block trackers. There's also increased adoption of the so-called ad blocking tools, which are generally open source and they use public filter lists to block ads and trackers on websites.

These ad blockers have become very popular. According to a recent estimate by PageFair, more than 600 million people around the world use ad blockers. According to a recent academic study by comScore, more than 18% of users in the US, use ad blocker.

In the male demographic, between the ages of 18 and 34, 50% of users in Germany use ad blockers. In the US this percentage is around 30%. So clearly, hundreds of millions of people are using ad blockers to protect their privacy.

Unfortunately, the online advertising industry sees these ad blocking tools as a growing threat. So they resorted to these so-called anti ad blocking techniques. So these websites employ these anti ad blocking scripts, which detect users who are using ad blockers. And then they force these users to disable their ad blocker or whitelist the website.

Many popular online publishers, such as the Washington Times, Wired, Forbes, have recently tried to interrupt and block ad block users. Such attempts to undermine ad blockers could mean a return to the status quo.

Therefore it is very important that we develop effective and long lasting countermeasures to circumvent anti ad blockers and strengthen ad blockers. So the goal of our project is twofold.

First we want to do a comprehensive measurement study to study how prevalent is this practice of anti ad blocking. So we want to answer questions like, how many websites are using anti ad blockers? Which popular third-party anti ad blockers are commonly used? And what kind of detection techniques do they use to detect ad block users?

And in the second part of this project, we want to improve ad blockers. We want to make a stealthy ad blocker, which would not be detectable by websites. So users can surf the web while making sure that they are not being tracked across the web, and they are not served malicious ads

So for the first part of this project we wanted to do a large scale automated study. To do this, we relied on this mechanism which we used in testing. So we opened up a website with an ad block and then without an ad block, and then we computed different content- based features.

And then we fed these features into standard machine learning algorithms to learn which websites are using anti ad blockers. And our models achieved more than 90% accuracy. And using this approach, we were able to detect more than 1000 websites out of top 100,000 websites.

So this study was done in early 2016, and these numbers have risen since then. So let me briefly talk about how these websites use anti ad blockers. How do these anti ad blockers work?

There are two basic approaches which are used by anti ad blockers. The first approach is to identify leaked extension information. So these ad blockers are generally installed as an extension or an add-on to the browser. So some of these JavaScripts can actually request for information, which can give them hints about whether a user has an ad blocker it or not.

Secondly, they use more intrusive techniques, which are active or passive, to detect whether an ad which was supposed to be there is currently also in place. So this is another technique using which they can try to tell whether users are using ad blockers or not.

So in the second part of this project, we want to make a stealthier ad blocker. And the current approach that we are investigating is, to develop a filter list of all these anti ad block scripts which these websites use. So if we can create a big filter list, we can remove all of these anti ad block scripts.

Right now, some of these efforts, for example, this anti ad block alert, which is very popular, it relies on crowdsourcing. And this is manually populated. So now the research project we are right now working on automatically populating these filter lists to make sure that these lists are updated very quickly.

And if websites quickly decide to change their tactics, we can quickly adapt and people can still use ad blockers effectively while browsing the web. So I would conclude with this quote by [INAUDIBLE] about ad block.

So just a quick show of hands in this audience. So how many of you actually use that an ad blocker? So that's good. That's more than half of the audience, so that's fantastic.

So I would like to really point out that ad blocking is not a problem, it's a symptom of a deeper problem in the online advertising ecosystem. Right now it's all about data. Companies want to money monetize user information, user data.

Right now you have to balance between users, society and economy. But right now these companies are putting the economy first and users at the very end. And this has to change. You have to put users first in this trade-off.

Our research aims to put users first. Our research is to make a stealthy ad blocker. We'll give users control over which ads and trackers are OK. So if they want to support a website, they can choose to whitelist a website and allow advertising on a particular website.

In the long term, there is a need to make better and more robust privacy preserving tools. Thank you.

KRISTIN KRAUSE COHEN: Thank you, Zubair. So we have a short period of time for some question and answer. And I encourage any of you in the audience who have questions to please come to the microphone, because we'd love to take audience questions. I'm going to start it off though, with James. I have a question about your research.

It seemed as if your paper concluded that because there was not a long term change in consumer sensitive search behavior that basically privacy choice was working. And I wondered if that was necessarily the case. Do consumers really have a choice in terms, if they want to stay connected what other choice do they have?

JAMES COOPER: Thanks. You know, I look at this like an economist. I mean, I'm a big tennis fan and I would love to go see the Australian Open in person, but I'm going to be forced to watch it on TV.

Because you know, it's expensive, and I have a lot of other things. So I'm not going to fly off to Australia this weekend. And I kind of look at it as the same thing. Now if Google changed their privacy policy, Google merges data here.

And if some people find it to the, and I measure maybe a 5% reduction at least in the month of March, if some people decide I'm going to kind of cut back on my sensitive search. I'm not going to do this because I'm a little worried about the stuff I've heard and I've read, then that's to me the market working.

I mean, and as I lay out in the paper, I mean it's kind of what you're measuring here is that to the extent that this reduction in anonymity or this intrusion in your autonomy deters you from engaging in behavior that you otherwise would, and that's what I'm measuring, then that's just like in any other market when the price of something goes up.

What happens is marginal people leave, and the people who were there maybe enjoy less, the info-marginal consumers enjoy less consumer surplus. But it's the workings of any market. If

price goes up, demand curve flipped down and that's what happens. And that's kind of how I view this.

KRISTIN KRAUSE COHEN: OK. Steve, I wanted to ask you about your research. What do you see as the next, what you hope to measure next with your tool. You did a lot of measurement and there was a lot discussed in your paper, but what do you see as the next area?

And for example, I noticed in your paper you looked at a lot of different kinds of canvasing, fingerprinting type of techniques. And this morning we heard from Aleksandra Korolova about the possibility of Bluetooth fingerprinting. Is that something that you would be able to look at with your tool? And what other areas would you want to look into?

STEVEN ENGLEHARDT: Yes sure. So I think one of the next steps coming is this notion of sensor based fingerprinting. You can imagine, if you want to re-identify or identify that two devices are in the same room and both of those devices were to have a rich set of sensors, you can imagine that the room is kind of identifying itself.

So if you're able to read information about the current status of a bunch of different sensors, you could know that both of these devices were in the same room. And that would be an interesting area to move into. I haven't looked into it at all.

But I think in other direction we're looking at is just kind of cross device tracking. We want to see how a company could do deterministic cross device tracking. Like where you can get access to a user's identity and use that to track them.

And the other thing is we want people to use the data. So please reach out to us and you can play around with our tool to get access to the data we already have. So that's like the second direction for us. Not so much the tool, but more on our data that we have.

KRISTIN KRAUSE COHEN: And do you do the crawl, I mean, I understand from your presentation, do you do it every month? And is that information always available via that?

STEVEN ENGLEHARDT: So we do the crawl every month. The data, everything I presented today is just from one month. That's all that's public right now. But once this tool is built and we actually flesh it out a little bit, we're planning to release everything as soon as it's collected.

KRISTIN KRAUSE COHEN: OK, great. Zubair, so in your presentation you mentioned that publishers are trying to monetize their content. Obviously, that's why they have these ads. But, you know, that they're so annoying and that's and the privacy implications. And that's why consumers have kind of flocked to these ad blockers.

But in your work, were you able to measure. I know in reading your paper that you were measuring how many anti ad blockers there were. And that also a great number of these sites it seemed like you could tell that they were checking for ad blockers.

And some of them were asking consumers to change the settings. Were you able to measure if they were successful in that? If they were able to show the value proposition to consumers of changing those settings?

ZUBAIR SHAFIQ: So in our study we were not able to measure that, because we were doing an active study. So we did not have data from actual users. But there have been anecdotal reports that users do actually sometimes respond to these messages. And some fraction of users do actually whitelist the site or turn off their ad blockers.

So there is another thing which has also becoming, which has become increasingly popular. These are these alternate revenue models. So websites are recognizing that it's not just ads that they can use to monetize their services. Some of these websites are asking users to subscribe or opt for a monthly donation.

So there are actually some exciting start-ups in this area about finding new and alternate ways to make revenue models for the online web industry, which replace the traditional online like targeted ads based monetization, which is obviously very intrusive.

KRISTIN KRAUSE COHEN: And do you know how successful those have been or are they just pretty, so new that there haven't been any studies?

ZUBAIR SHAFIQ: I know of two interesting ones. So there is a startup which is partially funded by Mozilla Foundation and it's called [INAUDIBLE]. They have their own browser. There is another effort by, I think it's Brave Browser. They also have their own browser.

And again, the goal of all of these new experiments and these start-ups is to give users more control. So they'll do this kind of ad blocking, but allow users an interface where they can decide which ads and trackers are OK.

And one of the interesting things they are actually doing is giving some of the revenue portion back to end users. So if a website makes some money off of it, they actually through this browser, gives some money back off the users for sharing this data.

So this value proposition is more explicit to users. They can see which websites are using their data. And then they are being paid by websites for sharing this data.

KRISTIN KRAUSE COHEN: That's very interesting

JAMES COOPER: Can I make one follow-up to that? Another thing I think is interesting about Brave is like, you'll see in some of these anti ad blocking messages like, oh, please sign up for our site. You know, $9.99 a month or $4.99 a month. And if you were to do that across the whole web, it's really, there's no way that the average user can afford that monthly charge per site.

So Brave is doing, I think it's implemented now, this kind of micropayment idea. So you can block ads on the site. It will let you block them, and then instead you'll pay for each page visit.

So it'll be interesting to see how that works out if users like it, if users are willing to pay per page visit instead of monthly fee for a specific site.

KRISTIN KRAUSE COHEN: Yes, we have an audience question.

BERIN SZOKA: Berin Szoka, Tech Freedom. It might be interesting, it might also be catastrophic. I mean you're talking about a model that has worked to deliver free content across the Internet, and that has allowed new start-ups and new media sources in the long tail to deploy and reach users.

So how do you think about, I know you can't study everything in one report, but how do you think about we could study in the future the effect on publishers, which in turn means the effect on revenue that's available for media outlets. That are dependent on the ad publishers for revenue.

And then the concentration. In other words, it might be that people are willing to pay for some things for the major sites. But there could be a very disparate impact where I might be willing to sign up for maybe just a handful of sites, or maybe only a handful of sites are going to subscribe to this new service that it was just hypothesized.

What makes you think that the web is still going to be as open and free and innovative as it has been before? And then finally, what about users who can't afford or don't have the technological sophistication to sign up for that sort of third-party service? When you start adding a layer of cost or hassle, how do we think about their welfare?

ZUBAIR SHAFIQ: So let me kind of jump in on this. So in my slide, which is also still up there, I think there are still three factors that you have to consider. Right now I think the starting point for any discussion should be that the status quo is not acceptable.

Right now companies are only optimizing economy. And the users and their privacy is not at the forefront of their thinking. So when you want to change this and you want to give some control back to users, so we have to work on some new models.

And some of these new subscription models, for example, the Brave One that was mentioned by Steve, they are fairly seamless. And they actually allow users to automatically without any hassle give some revenue back to all websites, including smaller websites.

And obviously, when this change happens, and I'm hoping that this change takes place, there will be new players in the ecosystem. And some players in that ecosystem would be fizzled out.

And I would like to think some of this would actually be good. It would be good if fake news websites are actually kicked out of the ecosystem. If these changes can actually help us achieve those kind of things, as well.

So I personally think it's not a win-win or win-lose. It's not a win-lose situation. Everyone can benefit. But we need to make sure that we put users at the forefront of this conversation on any new revenue model that gets widely used.

JAMES COOPER: I just want to, I guess maybe follow up with what Berin says. I mean, I wonder what the market failure is here. I mean, I'm an economist and I think, I think you're under the assumption that you say that we need to put consumers first, that consumers aren't first now.

I'm not really sure how, I guess I would ask what is the predicate for that assumption, that consumers aren't first in the online ecosystem right now? That, is it because the trade, the content that we're getting for free now isn't enough?

I mean should people have the right to say I don't feel like paying. It seems like you're trying to force a pay for content in something that's organically developed from an advertising, into an advertising. So I just wanted to know what the predicate for the market failure is here.

ZUBAIR SHAFIQ: So I think the fallacy of this argument is that you're assuming the web is free. It is not free. When you go to these services and you use them, you're actually implicitly paying them by sharing your data with these services. And data has inherent value.

So there is this payment that you are doing in terms of your data. So this transaction right now, right now it's only being optimized for these content publishers, but not for end users.

JAMES COOPER: When you say that, what do you mean it's only being optimized? So let's assume, like you said, the data is currency and I go and I, they get to track me and serve me ads. I'm getting something in return, too. Are you saying that exchange is not fair?

ZUBAIR SHAFIQ: So I'm not against like paying these services. These services obviously should be paid. But this value proposition should be more clear and users should have more control over what information they are sharing. So right now, obviously, it's so--

JAMES COOPER: I understand, but I mean, where, I know you're saying that they should have more control. Where does that come from? Where it is, I just wanted to know where are you saying that they have sort of lost control and that most people were unhappy with the way things are now and we need to have this change?

ZUBAIR SHAFIQ: I think most people are unhappy. The reason millions of people are using an ad blocker is a direct indication of that. So why would people install ad blocker?

JAMES COOPER: And what percentage of online users use an ad blocker?

ZUBAIR SHAFIQ: Increasingly greater percentage. In the US this number is around 30% and increasing. So it is a big percentage.

STEVEN ENGLEHARDT: I also just want to follow up on that and make the point that it's not necessary that we only need to move away from advertising and move to some kind of micro-

payment scheme. You could also move back to more contextual ads. Or even, there's a bunch of new research on like privacy preserving targeted ads.

So if you're sitting in the browser and making targeting decisions within the browser, there are ways to do that and report impressions and clicks and so on back to a service privately.

So it's not necessary that you have to move away from the free model. You could pursue one of these alternative models that aren't tied to data collection.

KRISTIN KRAUSE COHEN: I think we have another audience question.

Hi, Sarah Leonard. I work in ad operations. I manage a department at a publishing company. So I go to a lot of ad tech events and they have, a lot of the conversation is around ad blockers.

But in those conversations they talk more about latency and how people actually will implement ad blockers, not because they want to protect their data. It might be a positive thing, but it's because the latency involved with so many pixels firing onto the page is, cause such latency and we want our pages to load faster.

So a lot of that percentage of people in the US who, now I don't know about in Germany where they have a much higher percentage of use of blockers. I know that outside the US they're more, they're better about the use, about protecting privacy.

But here it might be because people just want their pages to load faster. So is it possible that they want this one thing, and they're also having a positive impact on protecting their privacy by protecting their data from being used on these pages. Your thoughts, because it was in your presentation.

ZUBAIR SHAFIQ: Right. So I think there are multiple reasons why a user may decide to install an ad blocker. So there are three big ones that I mentioned in my presentation, as well.

The first one that you mentioned is performance. So if you install an ad blocker you'll obviously have lower network traffic volume and the pages would load faster. But I think there are other benefits, as well.

So for example, the security benefit. We keep on hearing about malware are served by major publishers. This is another benefit that users get. And then the third benefit I think, which is one of the most important ones, is the privacy one.

So eventually it's a combination of these three factors which leads users to install ad blockers. So right now I'm not a aware of a user study being done to ascertain what is the kind of like distribution of users. How they weigh these different benefits of ad blockers.

But I think it's a nice package. And you get these three benefits in one and you protect your privacy, you improve your security, and at the same time improve your performance.

KRISTIN KRAUSE COHEN: We are running out of time, but I think we have time for maybe one, maybe two questions. Go ahead.

BRAD WELTMAN: Thank you, appreciate the research. I'm Brad Weltman with the Interactive Advertising Bureau and I wanted to make a very similar point, which is in many ways we agree with you. And we've been very public in saying that we have lost sight in some ways of the user experience, which is first and foremost what this needs to be about.

And all of the issues that were just raised, latency and battery strength, are what we need to focus on. And we believe, and actually some of our research shows, are really the main reasons that consumers are moving towards adopting ad blockers. Privacy is actually relatively low on the list.

So I wanted to just add that because I wanted you to give that full thought. I mean your Dracula quote sort of highlights that this is, there are three parts. And I think it's also important to recognize that there are different types of ad blockers.

There are ad blockers that are genuinely trying to provide users the experience that you're looking for them to [INAUDIBLE]. And there are others that have very different business models that are really leveraging the economic part of this industry in a very unfavorable way.

And to your point of you won't be able to pay for every service you can, I mean, I think that your research, I encourage you to keep track of that value exchange. Because that's really the central element here. We can control the parts as the online advertising industry that we can control and we're working to do so.

But the free content, the free services and the long tail that Berin pointed out, are an exceptionally important piece that I think need to be kept in mind as you continue to move forward on your research.

ZUBAIR SHAFIQ: I agree with your comments. But I just think that this value exchange has to be made more evident to users. And should favor users a little more in terms of their privacy, their security and also performance, which is one of the reasons people use ad blockers.

And people do care about privacy. For example, a large number of users use pure tracker blocking extensions like Ghostery. So for those people, the only incentive is to block these trackers. So I think this is a broader symptom in the online advertising ecosystem.

So it's good for the online advertising industry to recognize this problem and try to correct it, rather than trying to just look in the other direction or force users to uninstall ad blockers. That's not a good way to move forward.

KRISTIN KRAUSE COHEN: I would love to let everyone ask questions, but we are actually five minutes over. Would it be possible for you to come forward and talk individually with the panelist?

How about if we ask the questions and we can get their responses offline?

SPEAKER: I had a question for Professor Cooper. Running experiments in the wild is really complicated and I'm wondering how you took account for all the potentially confounding variables that could have influenced the results that you had?

Specifically, having done research, for example, looking at how the implementation of surveillance cameras might interact with crime patterns, one has to look at whether or not crime patterns change over time. So historically there's always a drop in March, for example.

One would potentially want to look not just at the search patterns on Google, but on Bing and on Yahoo and on other things to make sure that, in fact, you weren't just seeing a pattern that was happening globally. One might want to look at other things that were happening that might deter people from making a search request for a term like porn.

And so I'm just wondering if you could talk a little bit about how you're able to understand that something was not just a correlation, but there was actually some causal connection.

JAMES COOPER: I mean, since that was, can I respond publicly as opposed to--

KRISTIN KRAUSE COHEN: If you can do it in one minute.

JAMES COOPER: Yeah, I can. I used, I mean, I try to make causal claims. Like I said, it's a difference-in-differences. I have state fixed effects. I have week fixed effects. I have term fixed effects. I have term trends fixed effects.

I also look at different types of control, different windows of the control group to make sure that they're, I sort of look at the dummy interaction.

So I think, I mean this is all in my, I use a lot of econometrics, I'm not trying to use jargon, but I'm very aware of what it takes to make causal claims. And I think I control to the extent possible with everything I have.

Now using other websites or other search engines, I think that'd be great. They just don't have the data. And I am looking at the response to a Google privacy policy change. And so that's why I focused on Google.

I mean so so having data from other websites I think would be a fantastic control. I do do placebo checks, too. And like I said, I do do my randomized. Out of, in the distribution of 100 randomized runs that this regression leads me to believe the empirical distribution is kind of what I'm estimating with my point estimate.

KRISTIN KRAUSE COHEN: Thank you all so much. I'm sorry, Sebastian.

SEBASTIAN: I just have a very quick comment. Just for Zubair, I think what I'm always wondering is, if you consider the privacy policy to be a contract. Then what actually does this kind of self help lead to?

That's just one additional thought. And I know many people don't think of privacy policies that way, but might be worthwhile to do consider that as well.

KRISTIN KRAUSE COHEN: So we have a very short break. Please come back at 3:40. We have a great panel on information security and I'd like to thank our panelists for being here.