FTC PrivacyCon
January 14, 2016
Segment 4
Transcript

KEVIN MORIARTY: OK. Welcome back, everyone. My name's Kevin Moriarty. I'm with the Federal Trade Commission, and this is Session Four on the Economics of Privacy and security. First, we have Jens Grossklags from Penn State University presenting an empirical study of web vulnerability discovery ecosystems.

JENS GROSSKLAGS: So welcome to the first of two talks in the session that are actually about security. This is joint work with Mingyi Zhao and Peng Liu. We are all at the College of Information Science and Technology at Penn State University.

So my talk about a topic of bug bounties and vulnerability discovery that is mostly conducted by external researchers that are often called white hats. In 1995, the first bug bounty program was founded by Netscape that invited external security researchers who scrutinize its services.

Since then, we had a number of other company-sponsored programs emerging that were run in an independent fashion. However, more recently, we actually observed the merchants of so-called back bounty platforms-- two of them, HackerOne and Wooyun, which are the focus of our study. Wooyun was founded in 2010, and it's mostly focus on the Chinese market. HackerOne operates in Europe and in the United States mostly, and was founded in 2017.

So the motivation for our study is to better understand how these web vulnerability ecosystems actually operate, and whether they make a significant contribution to web security. We also want to provide useful data for the policy [INAUDIBLE], for example, on the limits of vulnerability research and practice.

Our approach is to do an in-depth empirical study of these two ecosystems, and in our paper, we take a very broad approach in the sense that we try to understand how organizations, white hats, black hats, the public interacts on these third-party vulnerability platforms. But in the presentations, I will mostly focus on the perspective of companies and organizations.

So the two programs that we look at have a couple of common aspects, mostly that they're very popular, and lots of white hats are interacting on them, and also a lot of vulnerability reports are made. But otherwise, there are a couple of important differences.

The first one is that HackerOne is organization-initiated in the sense that these companies ask HackerOne to run a particular problem for them, versus on Wooyun, hackers can actually submit-- any kind of hacker can submit to any type of vulnerability about any website to the platform.

So this is a very fundamental difference. There are some differences with respect to the bounties. Otherwise, also Wooyun is different in the sense that it has a delayed full disclosure policy. So

irrespective of the wishes of the company, after 45 days, the whole technical details of the discovered vulnerability will be communicated to the public.

So there are some differences in the type of data we have about the platform, so we cannot always directly contrast and compare the two. But what we can do is in five broad categories, provide somewhat of a comparison on how these platforms actually operate. The first one is participation. And what we observe here is that on HackerOne, the number of public programs that are run is limited to about 100. And all of those are IT companies.

In contrast, on Wooyun, we see a much broader portfolio of companies that are more or less coerced to participate on to platform. And interestingly, you see here, a lot of organizations are typically are not known to run bounty programs by themselves, like government institutions, education institutions, and also financial institutions.

So the first takeaway that we have is that the white hat-initiated model allows for a much broader participation, and which may be good in the sense of web security. The more limited participation model of platforms such as HackerOne, of course, raises then the question of how these platforms can actually encourage more companies to participate.

A second issue that you want to then explore is the quality of the submissions of what you observe here, in particular on the platform of Wooyun, is that you have a very broad range of types of vulnerabilities that are submitted. And in 44% of these cases, these are actually classified as high severity vulnerabilities.

On HackerOne, this a little bit harder to determine from publicly available data. However, if you actually peruse the bounty amounts of that are paid through the white hat hackers, and also look into the policy statements, by combining these two data points, we can actually also then infer how many vulnerabilities are of high and medium severity, which is plotted on the slide.

So here, we can also conclude that across these two programs, white hats actually make significant contributions to the security of these websites by contributing high severity vulnerabilities. But more broadly speaking, the white hat initiated model that we see on Wooyun seems to harvest more of these vulnerabilities in an efficient fashion.

Now the question arises how well actually these different platforms, and the particular companies associated with them, can actually respond to these submitted vulnerabilities. And here, we see some interesting differences. When we look at Wooyun, we actually that, in particular, those very popular companies as measured by Alexa rank, you see that most of them can actually adequately respond to the submitted vulnerabilities and handle them.

In contrast, less popular and smaller website very often are actually not capable to do so. So in fact, about 25% of the submitted vulnerabilities remain entirely unhandled by the organizations to which they are targeted.

On HackerOne, in contrast, since these are company-initiated programs, we see a very quick response time. Within four and a half hours, we see the first response to submit vulnerabilities,

and most of them are actually handled then within 30 days. So an interesting takeaway, then, is that on this white hat initiated model, on these platforms, we see that a lot of companies that are coerced to participate are actually also not prepared, which is something that we have to take in consideration.

And that, of course, raises the question of the balance on, well, should we actually coerce these companies to participate? Is it a reasonable activity that we should be engaged in? The next question that I'm approaching is, of course, [INAUDIBLE] interest, what impact do actually these kind of bounties have?

Here's a first overview that we are seeing. So what we are seeing here is a subsection or subsample of companies participating on HackerOne. We see on the left side that some companies are actually not paying any bounties at all, versus others pay actually pretty substantial bounties for a submitted vulnerability. On average, this doesn't really help us yet to determine what actually is a really significant impact is. And for that purpose, we actually conducted the regression analysis in which the dependent variable is the number of vulnerabilities submitted, and the independent variables are the average bounty paid by a particular program-- so the popularity of the program and the measure of the overall activity of the white hats on the platform in a particular period.

So what we are seeing here is-- first, I want to highlight the top part of the table, is that this about $100 increase in the expected bounty paid towards white hat researchers, we see about three more vulnerabilities reported to the programs. What we also see is that programs that are more popular are also receiving more vulnerability reports.

And that's, of course, has two factors. One, more popular websites, of course, we seek small attention. But often, they also more complex. They offer more services to the users, so they likely have a larger attack surface in the sense for white hat researchers to find potential vulnerabilities.

So the takeaway here is that white hats do not necessarily always focus on monetary compensation. In fact, what we observe is a 20% of all contributions on HackerOne actually go to those 33% programs that actually do not pay any bounties at all. So pay nothing actually serves as a potentially viable approach.

In contrast, what we also observe is, well, a higher bounty amount at the end of today is still associated also with a larger number of vulnerabilities that are submitted by the white hat researchers. So, which brings then to the last questions, the one about security improvements. So what do we actually get out of it?

And so in order to assess it, why we do not have an inside look into the organizations, we are using the trend of vulnerabilities submitted over time. So the argument here is that if you have a declining trend of vulnerabilities, everything else keeping moderately equal, then we would argue that perhaps at this particular website, the security's overall improving.

So when you take a first look at the data, then we see that it's actually rather spiky. So it's not immediately apparent by looking at these graphical depictions what kind of trends are emerging. However, one thing that we see in the top three graphs for the HackerOne is that there's this seemingly initial spike, where once the program is opened, a lot of vulnerability researchers are submitting vulnerabilities that they have stockpiled, or that they have been essentially energized by the opening of the program to do immediately a lot of research that led to additional submissions.

Wooyun is a bit more noisy. So in order to get a better understanding of how overall these trends shape up, we conduct a statistical test. It's called a Laplace Trend Test. And we focused here on programs that have a certain amount of minimum activity that we are running for at least four months, have at least 50 vulnerability reports submitted to them.

And what we see here is, actually, that there are two contrasting trends. So for HackerOne, we actually observe them over time for the majority of the programs, we see a decreasing trend of vulnerability reports. In contrast for Wooyun, which is a white hat initiated, coerced kind of company participation model, we see exactly the opposite, mostly an increase in the vulnerabilities reported.

So if we reason about that, then we could argue, well, despite monetary or perhaps because monetary incentives are in place, we actually see, never the less, these fewer vulnerabilities on HackerOne. So despite incentives, fewer vulnerabilities. We argue that this is indicative of actually improved web security practices at these participating companies.

And keep in mind, again, that these participating companies are mostly IT companies in the case of the public HackerOne programs. We also see this initial spike, which from a web security point of view, might be welcome news, if indeed it's indicative that a lot of stockpile vulnerabilities are actually removed from the knowledge of white hat and potentially black hat hackers.

We see an the opposing trend for Wooyun programs. And our interpretation of that is that, well, this likely has to do something with the lack of preparedness of these organizations when it comes to receiving these vulnerability reports.

So for example, they may not have a well-developed, secure, software developing life cycle. Good integration between the security team and these external security developers, and many other factors might actually play a role here.

Which already brings me to the last point. So we believe that it's instructive to conduct a really in-depth analysis of these programs to better understand what contribution they can actually make to its overall web security and practice. And it's definitely helpful that these two programs provided us public data, which we can study in detail.

There are many more results, which we actually have in the paper, in particular pertaining to how white hats actually behave. For example, we can showcase in our paper how white hats learn from one another by investigating the reports of their fellow hackers. We can also study what

kind of discovery patterns they actually have in place. For example, are they focusing on specific programs, or are they applying the same kind of technique across very different website?

So there are lots of interesting additional results, if you have already accumulated our papers, and I encourage you to take a look at them. In total, I believe that the jury is still out about which of these two participation models, the white hat initiated model or the company-initiated model, are really giving us the best advantages.

On the first glance, it seems that the white hat initiated model really has strong benefits in terms of participation. So we see many more whites hats, many more organizations that are involved in these kind of ecosystems. But on the other hand, a lot of these participating organizations are not very well-prepared when it comes to receiving these kind of vulnerability reports, and actually then improving also the security on their website.

So there is kind of pros and cons that we can observe. One issue that's clear is we can jumpstart and further engage in the discussion what kind of contributions overall these bounty programs make to the security office websites. Our initial assessment is positive, but I think we can go into further detail during the discussion, and that brings me to the end of my talk. Thank you very much.

[APPLAUSE]

KEVIN MORIARTY: Thank you. Thank you again. Next up, we have Veronica Marotta and Alessandro Acquisti from the Carnegie Mellon University.

ALESSANDRO ACQUISTI: Thank you, and good afternoon. This is a joint work. We have Veronica, [INAUDIBLE], and myself. If some of our previous work, you will know that often, we use behavior economics to try and understand how people make decisions about the personal information.

The study represented today is actually about traditional microeconomics. And it is about understanding the allocative and welfare impact of a targeted advertising. However, there is still a behavioral angle, at least in the motivations behind our work.

In the behavior decision research, it is very well-known that how you frame a certain problem influences the way people will think about this problem, and we make decisions about that. And currently, we live not only in the age of real data, but also under a very powerful frame, the frame that personal data is the new oil And we are all going to benefit, perhaps in the equal parts, from the collection and sharing [INAUDIBLE] of our personal information.

More specifically, there are a number although frames which are quite common in the family debate over privacy. For instance, personal information is the lifeblood of the internet. So the increasingly sophisticated collection of data is necessary for us to have free services online. Or lots of privacy is the price to pay to extract the benefits of the data. Or sharing personal information is an economic win-win, which benefits equally the data [INAUDIBLE] and data subjects.

Well, in our broader research agenda, we are interested in investigating all of these frames to see how actual empirical evidence there is supporting them or not supporting them. The paper we are presenting today tackles the last frame, and more specifically relates to the impact of the targeted advertising as on the surplus of different stakeholders-- consumers advertising firms, and intermediaries. The ad networks. And Veronica will guide you through the model.

VERONICA MEROTTA: So, thank you, Alessandro. So specifically, the question we are interested in addressing is, to what extent the availability of more and more precise information about consumers leads to an increasing total welfare, what Alessandro just referred to an economic win-win, versus a change of allocations benefits among the different stakeholders, including companies, consumers, and online intermediaries and platforms.

Now in order to address this question, we rely on economic modeling to be the most high-stage, [INAUDIBLE] model of aligned targeted advertising that compared different scenarios that differ in the type and the amount of consumer information that is available to the different players during the targeting process.

Now specifically, differently from previous work, we account for the important role played by the intermediary in the advertising ecosystem. And we focused on a specific mechanism of real-time bidding. Real-time bidding is a technology introduced to facilitate the allocation of programmatic advertisements online.

Let me explain to you quickly how it works. We have different players involved. On one side, we have publishers, namely websites, that wish to sell advertisement space that is available on their site. On the other side, we have advertisers, companies that wish to advertise their products online.

But those two players do not need to communicate directly. They can rely on the intermediary via the change that facilitates the location of the advertisements and the targeting process. So the mechanism works as follows. When a user arrives to a publisher's site, a signal is sent to the ad exchanged that's subsequently broadcasted along with user data-- maybe IP address, user cookies, geolocation-- to interested advertisers, and there's an aunction for the location of advertisement.

So on the basis of information that the advertisers receive, they form a bid-- so how much they're willing to pay to show an advertisement to the user-- and submit their bid to the ad exchange. Commonly, ad exchange uses second price auctions. This means that the highest bidder wins the auction, but he pays the second highest bid.

So once the winner is determined, he's allowed to show the advertisement to the user. Now, on of these mechanisms, we built a model that focuses on the interaction among three main players-- that advertisers, the intermediary, and the consumers.

We assume that the advertisers are profit-maximizing agents. They want to advertise their product to consumers that we like, and therefore buy their product. Nevertheless, they cannot

target consumers directly. They need to rely on an intermediary that facilitates the allocation's advertisements.

We assume that the intermediary itself is a profit-maximizing agent that receives a payment every time you hold the auction for the advertisement's allocation. Finally, consumers have product preferences, but they need to know which seller is selling which product. So in this sense, advertising plays an informative role to consumers.

Further, we assume that consumer can be capitalized by two categories of information-- horizontal information, capturing consumers' preferences and tastes, and vertical information, capturing differences in purchase power. Now, these two players interact in our model in this way. At the given point in time, a consumer is online, and he may be characterized by these two pieces of information, horizontal and vertical.

The other change received a signal about a consumer, sells the information, and holds an auction for the allocation of an advertisement to their consumer. On the basis of the information that you receive, advertisers form a their bid.

The auction is run, the winner is determined, and it is allowed to show advertisement to the consumer. The consumer sees the advertisement and makes this purchase decision. Now, it should be clear that the outcome of this process crucially depends on the information that's available during the targeting process. So therefore, we analyze how the outcome for consumers to advertisers and intermediary changes when the different types and amounts of a consumer's information are available.

We considered specifically four cases-- a case where only the horizontal information is available, a case where only the vertical information is available, a case were both pieces of information are available, and a benchmark case where no information about consumers is available. So, an extreme full privacy case.

For each of these cases, we arrive at what's the firm's best strategy, and therefore, what's the firm's profit, what's the intermediary received from the allocation of the advertisement, and what's the consumer's choice and surplus? Now, in the interest of time, I will not go through the mathematics of the model, but I would like to show you interesting results that we obtained by simulating the model.

So what we do, we run computational simulations to analyze the outcome in terms of consumer surplus, intermediary's profit, and advertiser's profit changes in the four different informational scenarios. Let me start from the consumers.

Now, the graph that you see here, the x-axis captures how heterogeneous consumers are in their preferences, while the y-axis captures how heterogeneous consumers are in their purchase power. Now, important to note, low values means a high heterogeneity. High values means high homogeneity.

Now, the different colors correspond to one of the different informational scenarios that we consider. Specifically, each region captures under which scenarios the consumers are better off. So we have two predominant colors here. The green regions captures all the combinations of the model parameters for which consumers are better off, when already the whole [INAUDIBLE] information is available during the targeting process.

So what's an intuition there? In their region, consumers are more heterogeneous in their product preferences. Therefore, revealing the horizontal information actually ensures the consumers see the advertisements for their products they like the most. So there is a better matching between consumers and companies.

The yellow region instead captures all the combinations of model parameters under which the consumers are better off when no information about them is revealed. So in their region, consumers tend to be more homogeneous, so brands don't matter as much, and so the targeting is not as valuable to consumers.

Now, we can construct a single graph for the intermediary's profit. Again, we have two main regions. The yellow region, again is the combination of model parameters for which the intermediary's profit now is highest when no information is revealed about the consumer.

So we said in the region, consumers tend to be more homogeneous. So what happens is that if advertiser had that information, they will tend to bid lower to show the advertisement, lower than the intermediary's profit. But if the information is not revealed, then the advertisers have to beat an expectation, so they may overbid, increasing the intermediary's profit.

The red region instead is the combination of model parameters for which the intermediary's profit is highest when the vertical information about the consumers is available. In the regions, consumers are more heterogeneous, and so revealing the vertical information during the targeting process intensifies the competition among the bidders. They may tend to bid more aggressively

So if we put together these two pictures, we see that we have situations in which the interest of these two players are actually aligned to the yellow region. But there are also situations in which they have contrasting interests. So we may think of a situation of an intermediary that may have power over the information about a consumer, and may decide to act strategically, either by revealing the wrong type of information, sea green versus red region, or revealing too much information, when instead consumers would've been better off with less information being revealed.

Now finally, we can use the simulations to understand and analyze how the allocation of the benefits among the different players changes under different scenarios. So we can construct a pie chart, like the one that we are seeing now for information case, where we see the percentage of the value generated to a targeting process that is captured by each period.

So we can have a pie chart for each scenario. And what these pie charts show is actually a pattern very similar to what we just discussed. Consumers in blue tend to be better off either in the no information case, on horizontal information case, while the intermediary in red seems to capture

a decent amount of the benefits in all the cases, with the vertical information one being by far the best case.

For firms instead intuitively, it's always better off to at least some of the information about the consumers, with the complete information case being in this case the best scenario. So if you want to summarize those findings, we find that consumers are generally better either when a specific type of information about them are available, or in general, when less information are available, and that there exist situations where the interest of the players, the intermediary and consumers, may be misaligned. And therefore, a selected intermediary may choose to selectively share consumer data in order to maximize its profits.

So I'll leave Alessandro to some final remarks.

ALESSANDRO ACQUISTI: Thank you. So there are a number of extensions we are planning on working on. Probably the most importantly is the empirical validation. In fact, if representatives of networks are in the room or following from our webcasts, if you want to disprove or prove our results, we would love to work with you.

Now, going back to the broader picture from where we started. On the left, you have the three frames I started from. And I claim that they have something in common, which has very little empirical validation. So I'm not claiming that they're necessarily wrong. I'm claiming that we really don't know how true they are. So on the right instead, I have three broad research questions that I believe are critical to really understand to what extent data is the new oil, and to what extent the benefits of this new oil are allocated fairly or not to the different stakeholders.

How is the surplus generated by data allocated? If we use privacy enhanced technologies to find a nice combination of protection of data and sharing of data, are there [INAUDIBLE] costs, and if so, who is suffering those costs? Individual consumers, because they may get less targeted advertising? Society as a whole, because maybe the next medical researcher investigating cancer cannot find the cure because he or she doesn't have enough data? Or is it just an issue of the decreasing the rent extracted by oligopolies in the data industry.

Very different scenarios. And therefore, also very different policy conclusions. And finally, under what conditions consumer do benefit from trades in their data, and under conditions they do not. Because I believe that the answer is not binary. It's not always good, or always bad. It is very much context dependent.

Now, this is work in progress. In fact, is work in our agenda. However, if you're interested in [INAUDIBLE] material in this area-- and by this area, I mean the economics of privacy-- you can find on SSRN a semi-final version of the paper that Curtis Taylor, Lea Wagman, and myself had accepted, and it's forthcoming in the Journal of Economic Literature. It's a view on the economics of privacy, and we will leave you with this. Thank you very much for your attention.

[APPLAUSE]

KEVIN MORIARTY: Thank you Veronica and Alessandro. Next is Catherine Tucker from MIT to present privacy protection, personalized medicine, and genetic test.

CATHERINE TUCKER: OK. Thank you very much for having me. So I'm Catherine Tucker, and I am an economist who studies the economic effects of different types of privacy regulation using real-life data. And what I'm going to be presenting today is joint work from Amalia Miller, where we investigate how different forms of privacy protections affect consumer take-up of genetic testing.

And because I know that a lot of you are here to think about advertising and more mainstream issues, I want to make a pitch for why this is interesting before you all go to your electronics. The first reason-- so why we think it was interesting is that, first of all, this is a technology with a huge upside, as I'll get to later.

Secondly, it's also a technology where I think even the most cynical person about privacy would say there are potential privacy consequences of this data being created. Sometimes, when you're thinking about targeted advertising, it's hard to actually articulate the privacy harm, which is why we often think about health and financial examples.

But if you think about genetic data, it's not hard to come up with examples of harm. So for example, I took a 23andMe test. I will share with you, I find out rather depressingly that I got a three times normal than average chance of getting macular degeneration later in life. That means I won't be able to see too well.

Now the reason I feel comfortable announcing it in this audience is because ultimately, I have tenure at MIT. I probably have the least potential consequences of anyone in the world of releasing that kind of data because I have a job and I have health insurance. But there are potential-- and you cannot have to go far to think of potential negative consequences of that data.

And as the previous presentation on genetic privacy articulated I think very well, there also issues to do with identifiable, the fact this data is persistent, and the fact that potentially, this data has spill over to family members. So it's really quite important privacy consequences.

The other reason I think this paper setting is useful is simply because there's been a lot of experimentation about different kinds of regulation, which allows us to have more of a horse race than we usually do when trying to evaluate how well privacy protections work.

Now, I said there's an upside to this day, and I just talked about the downside to it being created, but there's a huge upside. And the upside is the promise of personalized medicine. And the typical statement made in favor of the personalized medicine is that for the average drug, based on your genetic makeup, it won't work 25% of the time.

So we can imagine if we actually had genetic data, we'd be able to identify effective [INAUDIBLE], and save money on drugs, and I'd save money at the same time. Now, as well as these claims, I often find it useful to bring it to life with a very pertinent example, which is the example of Angelina Jolie, and her genetic test, and the action she took from it based on it.

So Angelina Jolie did genetic testing. She found out that she unfortunately had a mutation in her genes, which meant she was likely to get both breast and ovarian cancer, and as a result, had a double mastectomy and a hysterectomy.

Now, this is obviously a strident and decisive medical action. But in principle, it's going to reduce her chances of getting cancer by 70%. So this is the kind of data which actually leads to extreme forms of action in a medical sense, but there's a huge upsides in health outcomes in terms of it being create.

Now, what we're going to do in the study is look at state law's experimentation with different types of privacy regulation from 2000 and 2010. And what's nice about this variation is you always worry in any empirical study where the variation's coming from, why are the states actually experimenting in this way? Is there an underlying reason?

From what we can see, it was pretty random, driven by individual state senators who got a bee in their bonnet. And what also is nice is that they're experimenting with many different types of privacy regulation, and we're going to bucket them in the study into three, which are informed consent, regulating data use, and establishing property rights.

And I want to-- in the past, what I've done is I've said, well, the great thing about this is it actually emulates different country's approaches to doing privacy regulation, if you sort of think EU and OECD approaches, more associated with informed consent. Maybe the US, you could say we've thought about restricting data use.

And then, there's sort of this economy stream of establishing property rights. Now I've said that in the past, and the reason I no longer push it is I mentioned this once when I was giving this talk in Paris, and this person from the Ministry of Culture in France stood up and said, how dare you say that. In France, we regulate privacy in every single way you could possibly imagine. So it's not just one. But in general, what's nice about it is at least we've a horse race. For a different race, we might think about regulating privacy.

Now, we're going to have data on people's decisions to get these genetic tests. We're lucky we have a national sample that was done every five years in the period of studying, and they're going to be asking 30,000 people about whether or not they had a genetic test in each sample.

Now, it's a great data set in one way, in they focus on the decision to get a genetic test for working out whether or not you have genetic susceptibility towards breast and ovarian cancer. And the reason I say this is a very interesting genetic test is there's actually something you can do with this information to save your life if you take the test. So potentially, this is a hugely valuable piece of health data to create.

Now, the negative is that this is a technology in its early stages. And so as a result, we're only seeing a little bit of take-up in our sample, about less than 1%. Now what we're going to do in the paper is use standard econometric techniques to relate the decision of these people in our sample to go and get a genetic test to what the state privacy regime was like in that particular year.

Now, I realize this is not an economist audience, so what I want you to think of this is a statistical relationship that we do, where we're controlling for just about everything that you might think of going on in the background. We're controlling for the year, we're controlling for the state, we're controlling for everything about the patient.

Now if you like equations and subscripts, the paper's got plenty of those, so I direct you though. For now, for this audience, so what I decided to do is to present the main results in a bar chart. And the big punch line is that we bucket up our state regulations in this way, what we find is that when you have informed consent-- and that's informed consent where we're telling people how the data's going to be used, we get a reduction of third in terms of how many people are taking a genetic test.

Now, this is a large proportion, but remember, these are quite small numbers. So the baseline is small. Now when we have to usage restriction-- that is, we say or the state government says, this data can't be used to discriminate, say by employers, say by health insurance companies-- that really has no statistical effect that we can measure. The thing which has this big boost or positive effect on the decision to get a genetic test is whether or not you actually give individuals control over how that data will be used in the future.

Now, when you get results like this as an economist, you're always going to worry, well, where are they coming from, and what's the explanation? So one explanation which worried me was maybe it's not about the patients. Maybe it's about hospitals and whether or not they're offering the tests.

So we went to collecting more data to test this, and we found that's not really the explanation. It is the case that if you have these consent laws, hospitals react negatively. That's not a surprise. I found that in the past. Basically, it's because you have to construction an entire parallel system.

However, what was important about this study was we did find-- what we found was a negative reaction by hospitals in terms of whether they offered genetic tests to giving patients property rights. Again, maybe not surprising. Why would you set up a genetic testing facility at your hospital? Probably to do some research. And this is going to restrict your ability to research.

But it suggests that the main effect of having these individual controls positively affecting outcomes is not driven by the supply side, but instead driven by patients. Now, more proof of this-- this what, again, a typical thing we would do in economics is we're always going to worry about, well, you're saying it's about patients, but could be another explanation of something else going on in the state?

We tested for this by looking at all kinds of explanations. One such test was we looked to see, well, if we look at the decision to have an HIV test, which you might see-- think of it of similarly sensitive to having a genetic test-- could we see any influence of the genetic laws on that decision? We found absolutely nothing. We suggest it's not driven by underlying tastes for privacy in that state. Similarly, we couldn't find genetic law effects on flu shots, which suggests it's not driven by taste for preventative care.

So what is really going on? I've ruled out hospitals. I've ruled out just it being some spurious correlation to do with the state. And I think what we're going to argue is that ultimately, it makes sense when you understand how this privacy information is delivered.

Genetic testing is unusual in that you have genetic counseling where you sit down for genetic counselor, and you will discuss these privacy policies for perhaps 20 minutes, as well as the positive and negative consequences of taking a test. So this is very different from the typical online environment. We know that consumers actually find out about some of these laws.

The laws they don't find out about, though, are the anti-discrimination laws. These aren't usually part of the conversation. And I think that explains really the lack of effect. Consumers just aren't reassured, because they don't find out about these laws actually existing.

On the other hand, when you go through the typical forms or process, where someone is given informed consent, and told how the data could be used, but not correspondingly given control, we are going to argue that highlights a sense of powerlessness, which perhaps can explain some of the negative effect. Whereas when you restore control to the patient over how their data might be used in the future, then we have reception of control, potentially a positive effect, which might encourage them going ahead with the test.

Now, we have some more material in he paper, where we try and prove that this really is about privacy concerns, in that we show that these effects are going to be higher in situations where there's more likely to be bad news if you have genetic test, that is a reason to think you're going to have bad news from the test.

However, we also show there's absolutely no effect if you've already got bad news. That is, if you've already had cancer, the bad news is out there in your medical record, none of these privacy laws are actually going to drive any of the effects. I'm also going to show the effects are largest for people who, in their surveys, took various privacy protecting actions, such as refusing to state income. So again, that's going to draw it back to privacy, rather than something else explaining my results.

So let me just sum up what we found. So I want to emphasize, and I think this is important, with every empirical study, there's going to be limitations. And there certainly are in this study. We do our best to try and make it causal. However, you can always come up alternative explanations.

We don't actually see that in that patient and genetic counselor room, where they go through the privacy policies. So we're speculating on the mechanism based by reviewing the privacy policies we've seen in different states. And the other biggest advantage is that we're studying an early stage of diffusion, and so this is going to be representative of the individuals who embrace new technologies early.

Having said that, I think there is something to be learned, which is that when states give will control about how private information is shared, we do see an increase in genetic testing. And we see this increase particularly for people who are worried there may be bad news from the genetic test.

Now, we found that in general, informed consent-- that is, giving people information about how their data will be use, but without giving them corresponding control-- just deters patients, both patients and hospitals from having genetic tests and offering genetic tests.

Lastly, we found that data usage policies have absolutely really little effect. And so it's either good or bad news, depending on how you look at it. I was quite positively encouraged, because usually when I run a [INAUDIBLE] relationship between a privacy regulation and economic outcomes, I find a negative effect. So I was pleased to find nothing bad.

On the other hand, these laws are designed to help people, and perhaps my research suggests that they're actually just not being publicized enough to reassure patients. So with that, I will say thank you very much, and I thank you again to the organizers for giving me the chance to speak.

[APPLAUSE]

KEVIN MORIARTY: Thank you, Catherine. Next up is Sasha Romanoksy of Rand Corporation, presenting "Examining the Costs and Causes of Cyber Incidents.

SASHA ROMANOSKY: This It's been a long day, hasn't it? Thank you all for sticking around, and thank you to the FTC for hosting this. It's great to be here. I'm Sasha Romanosky. I will present some empirical work related to cyber events, and I'll define those in a second.

But I want to explain the bit of a motivation. There were these two motivations behind this work. One is, you probably heard of this executive order by the president a couple years ago to help improve critical infrastructure. As part of that, NIST developed this beautiful framework for cybersecurity. So if anyone has any questions about how to protect their systems, you can go to the standard and take a look, and it'll tell you everything you want to know.

The trouble with that is that it's a voluntary standard. It's certainly not meant to be regulated in any kind of way, despite some of the criticisms that people have had. And so the question then becomes, how do you get firms to adopt? How do you get firms to adopt these standards? We think they're under investing and security, so how do we get them to increase their security?

That's a great question. And so the story behind this empirical work is trying to understand the incentives of firms. Are there incentives? Do those incentives exist for them to adopt more security, or an appropriate amount of security, or a fair amount of security, an efficient amount of security?

We're going to look at. The other motivation-- for anyone who's had a conversation with me over the past few years knows that I am keen on cyber insurance, and the kind of empirical work that cyber insurance could perform, and how they can at the end of the day help assess the risk of firms. Really, that's what they're interested in, is understanding the variation in risk across their firms to price that, and the kind of de facto policy that they are creating now with these policies.

So with those motivations, what I look at-- the data set that I have comes from a company called Advisant, which is based in New York and provides loss and incident data to insurance

companies. They've been creating a data set on cyber events for a number of years now, but traditionally they look at loss of property, other kinds of general liability that firms will face. These are corporate data events related to loss and litigation.

Most of the data sets that you see up there relating to cyber events include 5,000, 6,000 observations. We have a data set of 12,000. So as far as I know, this is the largest data set of cyber events, data breaches, and privacy violations, which is very nice, because it allows us to do some analysis to try and understand better different kinds of patents and the risk that we'll talk about.

I'm separating the different kinds of events. When I say a cyber event, they're generally broke into these categories as I'm defining them. There's certainly other ways of categorizing them, and that's perfectly reasonable. For the purpose of my talk here, I'm separating them into data breaches, what we normally think about as an unauthorized disclosure of personal information, security incidents, attacks against a company for the purpose of causing harm to that company-- for example, a denial of service, or a theft of intellectual property, or an outage the system, and privacy violations. So, what I'm calling an unauthorized use or collection of personal information. And then, other sorts of phishing and skimming attacks. I think for this audience, we'll be mostly interested in the data breaches and the privacy violations.

One differentiator between the data breaches and the security incidents, what we might think of as acts caused to the firm-- so they are bearing. They are suffering these attacks, as opposed to privacy violation, where the firm is engaging in some kind of activity.

It's always useful to understand the data generating process, to understand where the data are coming from, and what's included, and what's not included. And so to be clear, these data come from public sources. There is no proprietary information.

Advisant has a wonderful team of analysts that go out and scour news sites, national and local news sites. Using the Freedom of Information Act requests, they find the information. Using Lexis, and Westlaw, and other data sources. So they've amassed this wonderful collection.

So a cyber event will occur to a firm, conditional in that it will be detected by the firm, either by the firm, by a third party, by a consumer, by law enforcement. Somehow, it's been observed by the firm. We, of course, have no information about those events which are not detected. That's just not in our data set.

Given detection, it is disclosed to the public. So certainly, of course, there's not always a requirement for a firm to disclose an event. There are exceptions, even with the breach notification laws. So we do not observe those that are not being disclosed.

Conditional on disclosure, we would hope it be would be reported within this data set. And of those events that are recorded in the data set, some will lead to a legal action, either private, public action, civil or criminal.

To give you a sense of the overall totals, we see that data breaches have been, in fact, increasing over the past few years. So these claims by others that there are more breaches now than there were before do seem to be true. However, we find that they're increasing at a decreasing rate, as opposed to security incidents, privacy violations, and these phishing, and skimming attacks, which represent a much smaller proportion of the overall incident.

So we see the first takeaway from this is that data breaches really represent the majority of these events. Interestingly, security incidents seem to be increasing at an increasing rate over the past few years. Now as far as I know, there have been no changes in regulation requiring disclosure-- an increase in disclosure of security incidents. And so, conditional on the same level of reporting and of detection, what this might suggest is that firms are being attacked more now than they were before.

In regard to the insurance industry in trying to understand the risk of their insured, one way to understand that is to look at analysis by industry. So we might want to understand what kinds of industries suffer the greatest number of attacks, or pose the greatest risk. And of course, there are many ways to think of this. We could look at total number of events by industry.

But that gives us an incomplete picture. And so we might look at the incident rate, the proportion, the percentage, of firms within a given industry that suffer the greatest number of attacks. And then, we could also look at lawsuits, just as an aggregate, and a litigation rate. We could also look at cost of events. I won't go through all of these in the interest of time.

But I'll show you the-- so as a function of total incidents, the finance and insurance industry suffer the greatest number of incidents, followed by health care and government, education, and then manufacturing. But as a function of incident rate, government agencies. So these are states and local DMVs-- law enforcement courts suffer the greatest incident rate, followed by education.

Let me just skip through these. And then we look at the legal actions. So of the 1,700 legal action that we have recorded in this database, 300 or so are criminal actions, some filed in the federal court, some filed in state court. But really, the bulk of these legal actions are private civil actions brought in federal court. And these will be allegations of-- all sorts of common law and statutory allegations. So negligence liability, and strict liability, a breach of contract, unjust enrichment-- a whole smattering. From previous years' research, we found over 80 unique causes of action brought by plaintiffs in these suits.

When we look at the litigation, the total number of litigation and the litigation rate, we see that privacy lawsuits have been increasing dramatically over the years, whereas the data breaches have been held steady. Now, specifically, these privacy violations-- in regard to the lawsuits, the privacy lawsuits, the allegations represent claims of typically, unsolicited email, or spam, or faxing, unsolicited telemarketing recording, either video or audio recording.

And overall, the litigation rate for data breaches and security incidents has been decreasing over the years, which confirms some of our previous work. And so right now, we're looking at a rate

of about 3% or 4%. What we also show here is that you'll notice the litigation rate for privacy violation is very quite high, 95%.

And I think this is really just more of an artifact of the data. I think while for the data breaches, we can understand a sample of breaches, and identify which of those have been litigated, because the breach notification laws. But for privacy violations, we don't really have that same denominator. We don't really understand the total number of violations, and therefore, the percentage of which would lead to litigation. I think in our data set, all we're really finding is we're only observing a privacy violation when a lawsuit is occurring.

Now the next question, we're going to look at some cost data. So I will couch this by saying that these are estimates of cost. They certainly don't include lots of other information. They're all firm-based. So typically, first party losses and third party losses. So all the costs that a firm would incur because of the data breach that you could imagine.

So the cost in notification, the costs of forensics, the cost of repairing any IT systems. In some cases, they represent a dollar figure loss, a financial loss. The third party losses the cost of litigating, of litigating the lawsuit of any kind of consumer redress, or financial sanctions imposed by regulating agencies.

So given all these costs, the big questions is, how much does a data breach cost? And so, Ponemon and others have produced some great surveys over the years, trying to estimate these costs. And what they come up with are typically figures of $5 million, $7 million, as the cost of a data breach.

I might argue, though, that this is an improper measure, because they're looking at the mean, the statistical average. And so because of the variation of the distribution of these costs, a median is a better metric. So not every data breach is a target of $270 million and rising. Not every breach is Sony. Not every beach is JP Morgan or Home Depot. There are many firms that don't lose that much money.

And so what we find here is that most companies lose less than $200,000. So if you were asking me the question of how much does a data breach cost, I would say less than $200,000. And this is getting back to the incentives that firms may or may not have in investing in security and privacy protection controls.

The median cost is a little bit higher for privacy violations. And that's still some work we're exploring to try and understand exactly why. But I think the takeaway here is that this $5 million, $7 million cost is overblown.

We also wanted to look at repeat players. So this notion comes up quite a bit in different conversations of, what is the impact to firms that suffer multiple kinds of events? Are they bearing high litigation rates? Are they bearing a higher cost? How often do they occur?

What we find is that in our data set, almost 40% of firms are these so-called repeat players, suffer multiple events. And that's quite a bit higher than I would have thought beforehead. I think that's

quite extraordinary, in fact. And indeed, in the information and financial insurance sectors, almost 50% of them are repeat players. I think that is quite interesting also.

The figures here that I'm showing, $9 and 1/2 million versus $4 are the mean. What it's showing you is that the cost for these repeat players is almost twice, a little over twice, than the non-repeat players, those that suffer just a single event. Now, the medians show exactly the same thing, but the cost is higher for these repeat players.

What I then also wanted to do is try and understand, OK, or maybe $200,000 is actually a lot for these firms, so what does this represent as a function of their revenue? So what I did is went through all of the data to try and understand, what do most companies lose as a function of the revenue, and then try and couch that relative to other kinds of losses in different sorts of industries.

So we wanted to look at retail. There's hospital, bad debt, global payment fraud. So what you could imagine is that Visa and MasterCard have a certain tolerance for risk or for fraud, for bad debt, and that through either an organic process or some calculations, they have settled on some percentage.

And these numbers come from industry reports showing 5.9%, 5.2%, 3.1% for fraud. Cyber events, less than half a percent. So it's saying that cyber events cost less than half a percent of a firm's revenue, a great deal less than these other industries.

In addition to that, in other works by some colleagues, Lilly Ablon at Rand, we conducted a survey using the American Life Panel, a great survey instrument the Rand has available to it to try and understand consumer sentiment towards breach notification.

How do they feel in response to firms getting these notices of a data breach? And what we find is that for the most part, they're really quite content. They're really quite happy with the responses that they're getting, with the timeliness, with the information presented in the notifications, and really have generally concerns.

There's a small percentage of them that may change firms, but by and large, they are really quite happy. So consumer sentiment, if it is in fact high, coupled with a small cost to a firm because of these events, really may suggest that firms have very little incentive to change their practices. Thank you very much.

[APPLAUSE]

KEVIN MORIARTY: Thank you, Sasha. And thank you to everyone for those presentations. They were wonderful, and very varied. So I want to recap them briefly. But first, I want to introduce Doug Smith, who's from the Federal Trade Commission, and Siona Listokin from the George Mason University School of Policy, Government, and International Affairs.

So we had four very different presentations. Jens presented an evaluation of two bug bounty programs, and offered conclusions about how they can be effective to identify, resolve, and

reduce vulnerabilities. Veronica and Alessandro proposed an economic model for advertisers, platforms, and consumers, and concluded that the allocation of the benefits of sharing consumer information tends to benefit the platform and the advertiser.

And if I'm wrong about any of these recaps, you can tell me in just a second. . Catherine presented an evaluation of the rate of genetic testing in states with privacy laws that fall into three different general categories, and concluded that states where redisclosure restricted have the highest testing rates, and that states with informed consent decreases the rate of genetic testing.

And finally, Sasha looked at one set of data and offered conclusions about the median cost of cyber events, putting it around 200,000 and less than what other studies have found about the cost of cyber events. So to start, I just want to turn it over to Siona to offer some thoughts and start the questions.

SIONA LISTOKIN: So Kevin had asked me to talk about themes in this panel. And I would note that the title of the panel is the Economics of Privacy and Security. And I think that's about as close as we'll get to a theme. Lots of variation here.

Papers covered some of the most important or touchstone topics in privacy-- so health data, online advertising, and of course, security. I'd also point out that the panel had a lot more focus on how firms respond to incentives, and not just consumers. And finally, a lot of talk-- the papers here really are a cross-section of stages of research design.

And so if you think about economics of privacy, we had a model that extends existing theory, descriptive papers using new data sets, and explanatory or causal papers. So my meta theme here is that the field of economics of privacy is alive and well and quite robust.

But that's going to be my question. So extending a Commissioner Brill's comments after lunch and the conclusion at the end of Veronica and Alessandro's paper, in this field, what's your wish list? And this is for everyone. Where do you see the gaps in this literature, specifically as it would relate to policymakers and industry practice? So not just advancing academic research. I'll start with Veronica and Alessandro, but I'm interested in everyone's thoughts.

ALESSANDRO ACQUISTI: One comment, and I'll piggy back on our last slides about the piece in JL, which came a lot from SSRN. In doing that review of the literature on the economic privacy, we identified three waves of research.

The field is not novel at all. It actually started in late 1970s, early 1980s, with Chicago School scholars, like Posner or Stiegler. So there is a beautiful pedigree and also, quite a bit of work starting back 40 years or so. However, only at the time-- there were no models, back in the late '70s, early '80s. No models or microeconomics in the field. Privacy was more about using economic concepts such as symmetric information, more or less out and apply them to privacy.

What we have now is lots of careful modern work. And what we started seeing in the last maybe five years, 10 years, thanks to the work of folks like Catherine Tucker and others, is beautiful

empirical work. So that is on my wish list, is to see even more empirical work. And in order to have more empirical work, sometimes we need data from the industry.

So if the industry is serious and believes really that data is the new oil, and that more transparency is good for everyone, we should have addressed the problem of informational symmetry, that surge was referring as one of the crucial problems we have in the previous panel, which is, we really want more data from the industry regarding exactly what they do with information they collect, so that even if people, the end users, may disregard the privacy policies, they may not care about what companies are doing, researchers can actually study the data, and then come out and aggregate it, and understand what is really happening, and then come up with maybe a policy recommendation. So my list is more empirical work and more transparency from the industry side.

SASHA ROMANOSKY: I mean, I would echo that, right? I think there's been a lot of time spent doing what a colleague would refer to as admiring the problem. And I think that's useful. And I think that's good. And I think that only gets us so far. I like empirical work because it speaks to evidence for something.

It gets us passed normatives, and values, and what should there be. And it really helps answer to the effect of, you know, what will be the effect of A on B. And certainly, the causal inference is the gold standard. And so in order to do that, the point is exactly true. We need the data, right? And sometimes that takes us being very creative on finding it in clever ways, like the previous panel the researchers did themselves, coming up with these experiments, which I think is beautiful.

And sometimes it takes pain for it, which is OK. But certainly, I think we need the empirical work. So I would echo everything Alessandro had to say, especially in the wonderful accented way that he said it.

CATHERINE TUCKER: Well, if I'll just to add to the accents. Unsurprisingly, I agree for the need for empirical work. What always strikes me is if I'm a policymaker trying to decide if I want a minimum wage, or what the level of the minimum wage should be, I could draw on hundreds of economic studies that have measured in hundreds of different ways how minimum wages affect wage levels.

However, if I'm a policymaker making really important decisions about whether to regulate privacy or data, I'm instead relying on just a handful of studies in potentially non-generalizable spheres. So really, it's almost personnel and numerosity.

JENS GROSSKLAGS: I want to add something on my wish list, is perhaps a better understanding of the long-term consequences of both the loss of privacy and the potential security compromises. And some work that Alessandro and I have done goes in that direction, to understand how people perceive privacy decision-making and what time, but what we could not assess in a robust manner is what are actually the potential losses that we may face down the road.

And I think this a very critical issue when it comes to genetic privacy, but also to consumer privacy. A similar issue also rises in the context of security, where actually the most interesting things might happen in the context of what we do not observe. And so you saw it in Sasha's chart.

We could only analyze the data that was detected. So what about all the security breaches that we do not observe and we know nothing about? Similar, with respect to my presentation, there is the behavior of white hats, which we can now analyze in a reasonable fashion, even though this was one of the first works doing that. But what we do not observe is the behavior of black hats.

And there, we still have a lots of work to be done in terms of investigating them, and getting maybe qualitative data, but also tying together data sets such Sasha's. This, for example, analysis that we have done to kind of be able to infer where vulnerabilities abilities have been known by the black hat community that have not been discovered by the white hats.

ALESSANDRO ACQUISTI: May I add something. Jens said something really important about long-term effects. And here is the dilemma that, as a field, economic privacy faces. In my belief, the most interesting applications of data sharing under the protection are long-term and indirect.

But generally, as economists, we can publish and do rigorous work when we have short-term and direct effects. It's very, very difficult to do studies and find [INAUDIBLE] links over long spans of time, when there could be a data breach now, which only has an effect the seven years later. And you are not going to satisfy reviewers in a legal historic journal with analyses that try to find those kind of effects. So these development [INAUDIBLE]. I don't think there is any simple methodological solution for that.

SIONA LISTOKIN: Thanks.

DOUG SMITH: I guess it's my turn. I have a question for the group, but I'm going to actually ask one for Catherine first. Catherine, what your research showed is that different laws have these different effects on consumers' choices in this particular context of genetic privacy. So I was curious sort of how you think this research-- what implications it has other areas of privacy and data security.

CATHERINE TUCKER: OK. So what was nice about this setting is it allowed us to have more of a horse race where we had the same thing we were trying to explain and lots of different privacy regimes. Now, the reason I find it useful or reassuring is it helps me believe some of the other research I've done in other areas that should be more case by case.

Some of the research I've done, for example, in targeted advertising, which a lot of people have talked about today, has emphasized the negative effects of informed consent, but also positive effects from improving consumer perceptions of control.

But I was always nervous, because those were two very separate studies, different at different times, different sphere, even different countries. And so I found it reassuring to actually use this

horse race to make me think, well, perhaps there is something more generalizable we can say about the effectiveness of different privacy regimes.

DOUG SMITH: Thanks. And then the question I have for the group is actually a little bit of a follow-up on one of the things Siona pointed out, which is, you guys are looking a lot at how firms' choices are happening in this arena. So what do these papers in general suggest about what the private sector is getting right, what it's getting wrong? What can this improve on our understanding of what kind of market failures you might be most concerned about in this area? Probably start this side, I guess.

JENS GROSSKLAGS: What's the private sector getting right? I think one observation also Alessandro and I have made over the time is that we see a lot of entities, private entities, entering the market, with privacy enhancing offers, but they're not really picked up in the marketplace to a sufficient degree. And well, the good news is that we do see these offers. We see a lot of technological solutions that are eventually picked up by start-ups.

But what we see less is an adoption by the big players because of the lack of incentives. Targeted marketing or advertisement is just too enticing to give it up in exchange for more privacy-friendly practical solution. So there's a fundamental conundrum that we are presented with that is very hard to sidestep. Nevertheless, I think it's very important that we see these new offers in the marketplace, and I hope more of them are actually picked up in practice.

SIONA LISTOKIN: What are they getting right?

ALESSANDRO ACQUISTI: Well, getting back to Jens' point about offers in the marketplace, one reason for optimism is the existence of privacy enhancing technologies, PET. So almost every time I'm invited here at the FTC, I end my talk about the PET, because I really strongly believe the technology is not just the problem, it can be the solution. But obviously, [INAUDIBLE] technologies do not stop altogether the flow of data, but rather modulate the sharing the protection.

So the reason for [INAUDIBLE] is that private sector firms can actually-- this may be wishful thinking, but may be proactive in deploying PETs, anticipating otherwise regulatory intervention so that they can still do much of what they're doing now, but in a more privacy-preserving manner. Now truth to be told, some of these technologies are still in their infancy. For instance, homomorphic encryption is very promising, but we still don't know how efficient and practical it will be. But the promising is enough for the moment, and I do believe that in the space of privacy, we can actually have the cake and eat it too, because of these technologies.

SASHA ROMANOSKY: In terms of what are firms getting right, god, that's such a good question. And I wish I had a better answer than the one I'm about to give. So I think what we can rely on is that firms will operate based on incentives. And of course, the goal, then, is to tweak the incentives such that they become aligned for all of the players.

Right? So that's not new. And what that means is that, look, if privacy really is a big deal, then consumers should really act like it's a big deal. And if and only until they do, will firms have

incentive to take it seriously. So I guess I would say that consumers should take it seriously, and act like it, then firms will take it seriously.

Now, if there are market failures for which consumers can't impose any kind of effect on the firm, then that's where regulation or policy or FTC actions could come into play. Go ahead.

CATHERINE TUCKER: No, I just want to build on that. Because I think what I often see in the discussion is this underlying assumption that it's never in the firm's interest to regulate on privacy. And therefore, government has to intervene. But I think there are instances that we see in research where there are incentives to firms to actually improve privacy protections for consumers-- for example, the provision of user-centric controls. And so, I sort of see that as a beam of light in an all too cynical world.

SASHA ROMANOSKY: Yeah. Yeah, and I think-- I mean, it does touch on the world of information disclosure, and choice and notice, and poor choice and notice. You know, poor choice. Over the past five, six years, it's taken a beating, hasn't it? But it's relied on this notion that firms don't behave the right way, consumers don't behave the right way because they don't have the right information. And only if we could give them the right information would they make the proper choices.

Ah, I'm just not sure that's true. At least let me say it this way, that maybe firms, at least in my case, with the data, that maybe firms do have the right information. Maybe they are aware of all the risks that using and collecting the data have, and that maybe they are making rational choices. And for them, investing a certain amount, which we may think his is underinvesting, isn't the proper amount. But maybe it is actually the right amount as far as they're concerned.

JENS GROSSKLAGS: I just want to also add that this panel was also about security. And I think one thing that firms do right is participating in bug bounty programs, and really taking serious efforts in hardening their web security, but also other security aspects. And I think they're still quite a step away from anything approaching full security, but I think having a multi-dimensional security program, including bug bounty programs, is definitely a step in the right direction.

KEVIN MORIARTY: Jens, on a related point, I wanted to ask you, there was a notorious blog post by the chief information officer of Oracle, where she sort of denigrated the value of bug bounty programs. And basically, the analysis was, look, it's very expensive to go through all these bug reports. It very rarely yields useful information.

Your study does show that there are benefits to participation. But I think the question that she was raising is, are the benefits of participation greater than the benefit of just using that same money to pay another engineer to evaluate internally the controls on your software?

JENS GROSSKLAGS: So bug bounty programs are certainly not the first security measurement that any kind of company should implement. However, as you saw on one of my early slides, actually, very mature companies from a security point of view, were the ones running their own bug bounty programs, like Facebook, Google, and so on.

So from their perspective, it was worth their while. And certainly, one of the main selling points is that it provides a different perspective in addition to running software office security tools, having internal security researchers, in the sense that white hat researchers have perhaps somewhat more of a view like a black hat in an organization. They are more creative. They poke holes in places where other the security researchers would not look.

And this is certainly a big selling point to inch the security of your website even a couple of steps further. Also, I think that's a lot of criticism about bet ratios between the reports and the data that is actually then useful. And I think when you actually look very closely at this, a lot to do with the matter of duplicate reports.

And well, I mean, this is actually white hat researchers doing their job. If the reports have not yet been disclosed, then well, they have will report oftentimes the same kind of security weaknesses to the particular entity, and well, taking this into account, then actually, the error rate is not that high.

My last point here is that here, actually, the involvement of bug bounty platforms can really have a positive impact, because they can introduce measures such as reputation mechanisms, coordinate [INAUDIBLE], and so on that actually then also instill some part of competition between the white hat community participants, so that they are more inclined to actually provide high quality data to the participating companies.

KEVIN MORIARTY: All. We have 20 seconds left. So Siona, do you have any final thoughts? Oh, Veronica, please.

VERONICA MAROTTA: Yeah. I wanted to clarify that in our findings, we don't find an intermediary always bad. But sometimes, it does do the right things for the consumer. So like, policy in this case is more nuanced. So there are cases in which the interest of the intermediaries is aligned with the consumers, but other cases in which instead, these incentives may be contrasting with the consumers.

KEVIN MORIARTY: And your paper is not currently up on our website, but I believe that we'll have it up following this presentation, so people can-- if they want more information about your findings, they can look at it there. Well, thank you all for participating in this session. We really appreciate it. Thank you.

[APPLAUSE]

[MUSIC PLAYING]