FTC Informational Injury Workshop
December 12, 2017
Segment 5: Panel 4
Transcript

DOUG SMITH: So hi. Good afternoon. My name's Doug Smith, and by co-moderator on this

GARRETT GLASGOW: Yes. Thank you for having me here. In terms of privacy and misuse of personal data, I'm working in two areas. One which I call the easy area is misuse of information by companies.

So where a customer and a company might have some kind of business relationship, certain levels of privacy or data protection are promised, and then the company, for whatever reason, doesn't live up to that, doesn't live up to the promises. I've been working on survey methods to

the negative [INAUDIBLE] they are imposing on the consumers. So in my view, that's inadequate.

In contrast, the ex-ante perspective would emphasize on the increase of risk to harm consumers, even if that risk has not being realized or could not be traced back to a particular firm. So I heard a lot of panelists talking about Equifax. Myself is a victim of Equifax. I worry about my record on the black market.

I end up paying something to try to freeze my account. I even sort of welcomed the inconvenience. I have to lock and unlock my account in order to get some credit in between. And to me, it spends my time and effort and money, and that is harm to me, even though we haven't seen the ultimate harm coming back in my way. And it's probably very hard to trace that back.

And similarly, if I know my favorite retailer has some bad data practice, even if that company has not had data breach, I will be worried. I would want to probably do something to reduce that risk. In that sense, the ex-ante perspective would say there's some harm there.

And to follow on Garrett's suit, we can make an environmental analogy here. So for example, a firm has polluted my neighborhood and exposed me to a higher risk of cancer. I would say the company should be responsible for that action, even though I have not developed my cancer in my body yet, because it increased the risk I'm exposed to. So in that sense, I believe the ex-ante perspective is more appropriate.

And then a point I want to make is that the biggest difficulty in measuring harm is not measuring harm itself. It's measuring harm attributable to a specific firm. We can measure the extent of identity theft or other things, but exactly how to tie that back to a particular firm, I think that's the most difficult question.

JACQUELINE CONNOR: Thank you, Ginger. Lynn?

LYNN LANGTON: Thank you for having me here today. As Jacqueline mentioned, I oversee the National Crime Victimization Survey, which is one of two measures of crime in the US. It dates back to the 1970s when it was developed to be a complement to police statistics because of the recognition that a large portion of crime goes unreported to police.

And so if we just focus on those crimes that are reported to law enforcement, we're missing a big piece of the picture. And this is particularly true when you're talking about sensitive crimes and when you're talking about more of what we would call sort of white collar crimes or emerging crimes, though I don't think emerging is really th 1lntarticoda2 (s)-esmu(a)82.2he o of.rvses, thwik4 (S)-2 (.id (

There is a lot of survey data in this space. A lot of sending out reports to different companies and

survey research. So maybe it was natural that I fell into a survey-based approach to looking at privacy and the value of data.

And there are two different approaches that I've looked at. There's what's known as conjoint analysis, and then there's contingent valuation. To different survey-type approaches. And why don't I just briefly talk about each then, the strengths and weaknesses of each of these approaches?

Conjoint analysis is well known in consumer research. It's accepted in court as a reliable method for uncovering truth in a lot of contexts. And what conjoint analysis involves is we ask our survey responders to make choices, as if they were in the market making choices from among some set of hypothetical products. And these products are going to have different features or attributes. And so we can see, if they make enough choices, which attributes they seem to value and which ones are unimportant.

I'll give you an example from my own research. We did research on streaming video services. And we present people-- we say, pretend you're in the market to update your streaming video service. Here are some services that might be available to you.

They might have different streaming speed. Some might have high definition available. Some might have more television shows. Of course, different monthly prices, and so on. And we just present these different services. We'd say, well, pretend these are the three that are available in your area or that you could possibly purchase. Which one of these services would you purchase? And they can make those choices.

Now when we want to bring this to privacy, one way we can do this is just regard privacy as just another feature of a product. Obviously, this is only going to work if we're looking at a situation where it's a consumer engaging in a market transaction with some company that's made some kind of promise about privacy or how they're going to treat your data. So we can say-- what we did with this paper that I'm talking about is part of that streaming video service was.

And then there's different possible privacy policies that these services offer. They might say, we never share your data. Others might say, well, we'll collect your viewing habits. Not your personal information, but we'll collect your viewing habits and package those up, and use those to help content providers decide what kinds of shows to make. And then a third option is, maybe we collect your personal information as well, and we might share that with third-party marketers, and so on.

And what we were able to do with this survey was see, what value do people place on protecting their data? Or conversely, what kind of discount do we need to offer consumers to get them to

I mean, I suppose we could design some kind of conjoint analysis. We'll say, now pretend there's a certain percent chance that this company's going have a data breach in the next year. But I don't think most consumers think about their purchases that way. It's a strange hypothetical to pose to people, and I think you'd get really strange results with a survey like that. Maybe it's possible to do, but I think that's one of the main weaknesses.

But at least in areas where we can apply this, this conjoint analysis lets us measure the value that individuals are placing on their privacy or on different data sharing policies. It gives us what we call a willingness to pay to protect their data.

And then we're talking about quantifying damages. If a company doesn't live up to that promise, we can use that willingness to pay to calculate damages, that we can use that as the base just to say, a certain percentage of this price was privacy protection. You didn't live up to privacy protection. You owe consumers some refund or some amount of damages based on your failure to live up to your promise.

And that's-- I mentioned earlier, there's easy cases and hard cases. I think those are kind of the easy cases. I think we've made some headway in terms of measuring the value of information, the value of privacy there.

Hard cases are all the other ones, things like data breaches, and so on. And one possible approach to that is what's known as contingent valuation. And a

The strengths here are that unlike some other surveys, this is something we can at least apply to something that's not a traditional market transaction. A data breach is not a traditional market transaction.

And what we found was that when there was no cheese pizza involved, people tended to behave in a way which was consistent with what they said in a survey about their privacy preferences. But the moment we gave them the offer of cheese pizza, even those people who said they really cared deeply about privacy, started giving away this quite sensitive data.

Now in terms of interpreting it, it's really quite hard, right? One thing you could do is sort of take an economist response, which is that stated preferences, it's hard to use them to measure anything. We should use revealed preferences.

Another interpretation of the study is, oh my gosh. If MIT undergraduates behave like this, people who should really understand technology, maybe we need to really protect people to actually get them to behave in the line of their stated preferences. So no conclusions. Just some difficulties in terms of using survey responses.

DOUG SMITH: Thank you, Catherine. Ginger, do you have any thoughts about--

GINGER JIN: Yeah, I would just want to follow up with Catherine's commons on the privacy paradox, and how actions seem to differ a lot from stated preferences. I think the biggest question is to distinguish different explanations behind this, because consumers don't know the risk, so they're willing to give it away, or because they know the risk, but they somehow believe the benefits dominate the risk.

Or they see the extra risk of giving away this to one more person, given that their data probably already been breached multiple times is so small that they feel helpless, and so therefore, they give away. I think those explanations would sort e

happening. That's only 5% of the residents that are participating in the survey that say they actively do this.

So now I'm thinking, what could we ask? Could we ask some questions related to data breaches specifically that would get at the behaviors that they engage in? So even just, how often do you consider whether a company has had any sort of breach that you can find information about before you provide your information to them? Just to kind of gauge whether or not there are actions being taken to actively avoid potential situations.

And again, that companies can change over time, and so those are challenging things to measure. But I think there are still some actions that we could think about concretely getting at that would address this issue a little bit more.

DOUG SMITH: Thanks, Lynn. And I think with talking about that, you sort of broadened the conversation to looking directly at people's actions, or at least their reported actions. So Josephine, would you like to comment on either sort of stated preferences or revealed preferences, or both?

JOSEPHINE WOLFF: Sure. Well, I think that a lot of what we have when we look at the data that people have analyzed and collected around the injuries and the costs associated with data breaches is very much stated preferences. Most of what you see, especially say in industry reports, is going to be kind of self-reporting, this is how much this breach cost, or this is how I

And that kind of decision, I think, is not actually as straightforward as saying, oh, well, I guess nobody cares about the injury that was done to them in this data breach because they're not willing to pursue this lawsuit. But more about trying to understand, what are the actual decisions that people are making when it comes down to acting on what they think are their preferences, on what they think is what they value?

DOUG SMITH: So in some sense, understanding sort of the costs that they're trading off, the benefits of privacy is sort of important to figure out how to measure it?

JOSEPHINE WOLFF: Absolutely. And I think gets at some of that discrepancy we see between sort of how much I think I value my privacy and the things I'm actually maybe willing to do in practice to protect it or pursue my losses.

DOUG SMITH: OK, great. Thanks, Josephine. Yeah. So do people have any further thoughts about sort of measuring consumer preferences through their actions? Garrett?

GARRETT GLASGOW: Yeah, actually. I think we've heard several different stories now of a mismatch between consumer stated preferences and consumer actions. And this is a regular feature of privacy research. We see this a lot, that people seem to take actions that contradict what they've stated matters to them.

So I certainly think it's true that the self-reported data has a role to play in this. I think where it becomes tricky is when you're looking at really kind of crisp quantification for things like legal remedies or policy interventions, whether there's any way to turn that into something that can be calculated precisely enough for those types of remedies.

DOUG SMITH: Thanks, Josephine. OK. So we're going to have to leave the topic of preferences and move on to an alternative, which is to just straight out measure outcomes and try to understand things that way. So Lynn, I'd like to hear more about how your study allows us to get at that question.

LYNN LANGTON: So again, we're talking in the context of identity theft here. So beyond just any sort of data breach or privacy-related issue. But we ask a whole series of questions related to both tangible and intangible harms. Again, ex-post harms associated with the misuse of personal information.

So of course, we ask about financial losses. That's an obvious one that you have to include. We also ask about the amount of time that individual had to spend clearing up the issues related to the victimization. And then we ask a whole series of other questions, trying to get more at some of these intangibles.

So I mentioned already that we ask about distress. We also ask about whether the incident resulted in any problems with family, friends, work, school. And then we ask questions about whether they experienced any credit problems related to the incident, whether they experienced any legal problems related to the incident, whether they had to deal with debt collectors. So these are some of those more intangibles.

And then I think to get more at those intangibles too and really the impact of an incident on an individual-- and again, this goes broader than identity theft. You know, the other thing that we can do is then sort of cross these different types of outcomes and different types of harm.

So when we look at, for example, how much time an individual has to spend dealing with an issue and then we look at the level of distress, I mean, there's a positive linear relationship there that's pretty strong. So when an individual has to spend six months or more dealing with this misuse of their information, a large portion of them say they were severely distressed. Whereas if they're able to resolve it in a day or so, you know, a smaller portion say that it was severely distressing.

So again, it's indirectly translatable to data breaches, but I think sort of the same ideas. And you can use survey research to sort of tap into these ex-post responses and harms that victims experience as a result of these incidents.

DOUG SMITH: Thanks, Lynn. Catherine, do you have any thoughts about this? About sort of measuring outcomes, what data we might look for, or how we might take things that are sort of less tangible and convert them into something we can sort of quantify?

A

So that, I think, is the hard case, once we introduce this uncertainty. And we can measure things

[LAUGH]

--ownership, maybe that technology could be used here to track sort of how the data changed from one hand to the other. I'm not a computer scientist. I don't know exactly how to do that. But I have seen other people, both computer scientists and economists sort of try to work in this area. To me, that seems a pretty promising direction to really do some research on.

DOUG SMITH: Thanks, Ginger. Lynn.

LYNN LANGTON: So both Ginger and Josephine made the point that causal ordering is really difficult to establish when you're talking about individual incidents, and I would certainly echo that. And it's something that we've wrestled with quite a bit with the supplement. Are there ways that we can more directly try to tie experiences of data breaches to individual incidents of identity theft?

And the reality is, I mean, when you're collecting survey data, the data are only as good as the responses you get. And we know, because we already asked respondents if they know how their information was obtained, that the majority of victims don't have any idea. So we have about 30% of our victims that say they have some idea, even if they're not sure, about how their information was obtained.

So using a survey like the NCVS to try to get at this causal relationship between data breaches and identity theft is really not the best vehicle, unfortunately. I think you have to look for sort of these other technological, computer science-based solutions.

I mean, among those that do say they know how their information was obtained, about 20% of our victims say that it was obtained through personnel or other files, personal information being obtained through a company that had their information. But again, that's such a small percentage of the 30% that knew how their information was taken that we can't really use that to draw any conclusions. The causal ordering issue is a big problem for that.

DOUG SMITH: Thanks, Lynn. Catherine.

CATHERINE TUCKER: Oh, no. I think, you know, Ginger and I completely actually agoetn tthes we es we tere

JOSEPHINE WOLFF: I would just say, I think the flip side of that, which is also an important area for more research both from academia and from government and industry is, what prevents harm from happening? Which I think is also a pretty underdeveloped area.

And that when you shift the question and you say, OK, what is it that we think people are actually willing to spend money on, it turns out there are a lot of different kinds of harm or injury that people might be interested in trying to insure themselves against. And that at least some people are willing to spend money on, which I think does give you a sense of what people really care about and want to protect themselves against. But again, at this moment, we're talking about a fairly small population.

JACQUELINE CONNOR: OK, great. And we have another question. And maybe, Catherine, given that you opened up talking about data breach notification laws, you might want to take a first stab at it. But the question is, can panelists discuss the multitude of state breach reporting rules and how that complicates setting a national standard of harm or injury?

CATHERINE TUCKER: Well, I'll just start off by saying, so as part of this study we did about hospital data breaches and data breach notification rules, we spent a lot of time trying to decode the different texts of the laws, and there's an amazing amount of variation. And the other thing I would say is that in my very lay opinion, many of the laws seem rather inconsistent if you know anything about technology in there were exemptions which make no sense, and it looks like, I don't know, people were just taking a random word generator sometimes to actually try and describe what they wanted to happen.

Now maybe this is an opportunity for something better. But certainly, I had a certain amount of disquiet, having seen the lack of standardization of language in these laws. And I'm an economist, not a lawyer. So that's probably a bad thing.

JACQUELINE CONNOR: Ginger?

GINGER JIN: Can I add? I think given the discrepancy we see across states, there's definitely a value to standardize the notification law, just to make sure that firms know what to do after a data breach.

But I want to ask probably a harder question, that we got a concern about what happened after notification. If the firms sort of meet all the obligations stated in the law and have disclosed the information they know at that moment. So what? We're relying on the consumers or the media or the public somehow respond to it so that they would feel embarrassed and therefore improve their data security? I mean, to me, that sounds like pretty wishful thinking.

As we know, if the harms we have in mind cannot be traced back to the individual firms, and they've already done their duty in notification, it's almost like, OK. I have done what I can, right? The rest is up to you. It's up to the consumers and vigilance of sort of having their own preventive measures or other things. I think that question has got to be coupled about the data notification law itself.

JACQUELINE CONNOR: Does anyone else have any thoughts on that question?

GARRETT GLASGOW: Just on reporting requirements in general, say, you know, of course we want-- if companies lose control of our data, of course, consumers need to know about that. But I

GINGER JIN: Well, one thing we haven't touched much in this panel, but has been touched in the previous panel is some similarity between the problem of privacy data security and the problem we have seen before in, say, food safety, drug safety, product liability, and [INAUDIBLE] laws. And so I would like to see probably more interdisciplinary research to sort of summarize the lessons we have learned from those areas and to see to what extent we can sort of apply the insights we have learned those to the market of privacy and data security.

GARRETT GLASGOW: And I think an important topic that isn't well understood yet that we should push forward on is maybe a theoretical or definitional issue of what we mean by informational injury. Does informational injury spring from the content of the information itself? Or is it how it's treated by, say, a company that you're doing business with?

And so here's a thought experiment. Suppose you're doing business with a company and they have a data breach, and some of your personal information is stolen and is now out there on the black market. But then you find out the previous week, there was another data breach with a different company, and all that same information was already out there a week ago. Were you harmed by that second data breach or not?

And whether or not you were harmed might depend on what you think of is informational injury. Is that the fact that this information is now out there? If that's the harm, the harm's already done. Or is it, this company is mistreating customer data and is not being fair to customers, it's dealing unfairly? Maybe that's a different kind of informational injury, and in that case, you could be harmed twice by the leak of the same information.

Both of those things could be sources of harm. It's entirely valid to believe they're both sources of harm. But I think that's something that is important to distinguish when we think about what kinds of enforcement we want to do and what kinds of harm we're trying to measure.