

FTC Big Data: A Tool for Inclusion or Exclusion? Workshop
September 15, 2014
Segment 6
Transcript

CHRISTOPHER OLSEN: Thanks everyone. Thanks everyone for joining us for the final panel. We're here talking about big data, and a lot of people talking about leaving digital footprints-- somebody left physical evidence of their person in the ladies' room, so reading glasses, they're up here to be claimed. And I know there's no coffee allowed in here, which I think is sort of a disaster for the last panel of the day. So we're going to try and--

SPEAKER 1: [INAUDIBLE] break at a quarter of. It closes at 3:00.

CHRISTOPHER OLSEN: Closed at 3:00. Yeah. We screwed everything up, didn't we? We have screwed everything up and we haven't even started.

SPEAKER 2: I blame the FTC.

CHRISTOPHER OLSEN: All right, this panel is on paths forward. I have a very distinguished group of panelists here with me. It's going to be a challenge for all of us because a number of panelists earlier in the day discussed steps forward. So this panel is challenged to come up with something new and different for the last panel. But I'm sure they're up to the task. Just quick introductions, and I should have borrowed the reading glasses that I just had.

To my left, Chris Calabrese is the Legislative Council for privacy-related issues in the ACLU's Washington office, where his portfolio includes internet privacy new surveillance technologies.

Next to him, Dan Castro is a senior analyst at the Information Technology and Innovation Foundation and the Director of the Center for Data Innovation.

Jeanette Fitzgerald, next to Dan, is General Counsel and Chief Privacy Officer for Epsilon, where she leads the government affairs, legislative, and regulatory initiatives related to data protection and privacy.

Jeremy Gillula-- did I pronounce that right?

JEREMY GILLULA: Yup. You got it.

CHRISTOPHER OLSEN: All right, is a staff technologist at EFF, the Electronic Frontier Foundation, where he focuses on privacy and civil liberties issues arising from new technology.

Next to Jeremy is Michael Spadea, a director at Promontory Financial Group, where he advises clients on a wide range of regulatory and compliance issues related to privacy and information governance.

And last but not least, Chris Wolf is a senior partner at Hogan Lovells, where he leads the firm's global privacy and information management practice, also the founder and chair of the Future Privacy Forum and chair of the Anti-Defamation League National Civil Rights Committee.

So to kick us off, I want to do something a little bit different. And I didn't warn the panelists about this in advance, so this is a classic moderator foul. But I'm going to proceed anyway.

CHRIS WOLF: We were told there'd be no quizzes.

CHRISTOPHER OLSEN: So I'm going to start with a McLaughlin Group-style question. There's been a lot of discussion today about practices that are occurring and could occur, and there's been discussion about the legal landscape and the regulatory landscape. And I'd like to ask each panelist for a yes or no answer to the following question. You can say it depends, but that's really cheating, so I wouldn't go with that. Do you agree that there are currently uses of data or potential uses that are harmful that are not addressed by the current legal or regulatory landscape? Chris.

CHRIS CALABRESE: Yes.

CHRISTOPHER OLSEN: Dan.

DAN CASTRO: I don't think we've heard any today.

JEANETTE FITZGERALD: No.

JEREMY GILLULA: Definitely.

MICHAEL SPADEA: Gun to my head, no.

CHRIS WOLF: So I'm a former litigator, and I would never let a witness answer a yes or no question that needs explanation. So we'll be discussing this.

[LAUGHTER]

CHRISTOPHER OLSEN: OK. It sounds like we've got a mix on the panel. And I think, before we get too much into specifics about how we might move forward, it might behoove us to flesh out a little bit of the answers that have been given to the simple question there. And I would ask each panelist to talk about whether there are legal gaps or market failures that are not being addressed in the first instance. And I'll just start with Chris.

CHRIS CALABRESE: Sure. So just to give a frame for this, data is not bad. It's not good, either. It just is. It's a fact of the environment. So it reflects existing disparities in our society. We see a lot of money in this country that is distributed along racial lines. And so we are going to see those distinctions. I believe the wealth gap in this country, white households now have approximately 20 times the average household wealth of black households. So the data is going

to reflect that. So our job here is to make sure that big data does not exacerbate it, and then ideally, hopefully down the road, can help to close it. But let's start by not exacerbating it.

So, potential regulatory gaps. I am very comfortable saying that there are regulatory gaps. And I'll give you a couple. One of the major ways that "big data" and data is combined today is in background checks. So if you want to see whether somebody's got a criminal background or not-- and I know it's covered-- is you do this background check. Lots of public data sources are checked. And we say, Chris Calabrese, criminal record in Texas. That is not true, by the way. And so Chris Calabrese probably doesn't get a job.

Well, we've seen lots and lots of examples where there are multiple Chris Calabreses, and there are, and those are mixed up. Well, I see a great and classic example for a market failure here, because the customer is not Chris Calabrese. The customer is the company. And he or she, it, may be willing to deal with a certain level of error if it improves their bottom line, which is not to hire somebody with a criminal record. They may be willing to accept a certain amount of data problem in order to deal with that larger problem.

Similarly, I have a product that detects fraud. If I'm a big bank, I'm really excited if I cut my fraud down by 40%. If I have 2% or 3% of people aren't able to get a product or have to jump through more hoops to get a product, that's fine, because that's not really what I'm worried about. My desire is to reduce fraud, and I'm willing to accept a certain amount of error to do that. And some people don't get products, that's too bad. But again, the market isn't going to fix that. So I'll just leave with those two.

CHRISTOPHER OLSEN: How about you, Dan?

DAN CASTRO: Yeah. So I think what's really interesting about conversations like we've had today, which has been very productive, is we have a chance to have a lot of voices in the room share where they do think there are problems. And so listening to the discussion today, and that was to my answer, I didn't hear a lot of real specifics about where there was something that wasn't being addressed today, where somebody was standing up saying, look, this is how I'm being harmed today, and this is the reason nobody can take an action.

And I think that's what we have to talk about when we talk about regulatory gaps. It's not enough to say there might be a problem. And the reason this matters is because there's so many opportunities to use big data for good. And that was part of what the first panel talked about. And so when we're talking about regulatory actions, we know there can be unintended consequences. There's always unintended consequences with any action. And so we have to be asking, in this case the FTC, make a good cost-benefit analysis of any type of proposed action. Any type of proposed intervention, you have to know what the costs are. You have to know what the savings will be.

But just to also pick up on something, Chris, you had said. You had said, data just is. And I guess it depends on what your definition of "is" is here. But I think that's actually the wrong approach. And we just came off the legal panel, so I'll throw it back to Lawrence Lessig and his famous line about code is law. In this case, data is law, and data isn't natural. It's something that's

created. We have to think about how it's created and the implications of its creation. And part of us doing that helps solve some of these types of problems that we confront. And that's not a regulatory solution. That's a technology solution.

CHRISTOPHER OLSEN: OK. I want to come back to you, Dan, a little bit later and talk to you about how that squares with what you've written about in terms of the data divide and the concerns about collecting data from sources that may not be equally available to all particular groups, and whether that presents a problem, maybe not necessarily from a regulatory perspective but maybe from a policy-making perspective. So hold on to that thought for the future. Jeanette, how about you?

JEANETTE FITZGERALD: Sure. But first off, I'm glad you're coming back to him because I wanted to hear more of what he was going to say too. I wish he'd kept going.

So I said no because I think there's a lot of self-regulation that already exists. The DMA is out there, the IEB-- there's several other of those A's and B's and all those other groups that have all self-regulatory guidelines. I know for the DMA, they will enforce those guidelines among their members. If they hear about somebody who's not a member, they would go talk to them and try and get them to act in what is considered an ethical manner among that group. And if they then still find that there's a problem, they've been known to turn those companies over to the FTC so they can look at them further.

So if there's a problem that somebody thinks is in a gap, then maybe we can address it that way without having to come up with another law that will only deal with a certain or broad-ranging area. But it won't get to what the real problem is, because as you said, all I've heard, and in all those reports that came out, they said it's possible there could be a harm. It's possible, but I couldn't find one either.

CHRIS CALABRESE: I have more.

JEANETTE FITZGERALD: OK. Good.

CHRISTOPHER OLSEN: How about you, Jeremy?

JEREMY GILLULA: So approaching this from sort of a technologist perspective-- I mean, I said yes because, just thinking about it from a statistical perspective, if you're trying to classify something and you get a 97% percent success rate, that's amazing. That's what people get tenure for if you can pull that off. And that means you've still got 3% that are wrong.

If you're talking about classifying every person in this country, that means you're wrong six million times or more if you've got a 2% or 3% error rate. And that's a lot of people that your automated decision-making based on big data could be harming. And I think that until we actually-- it's a different thing when you're doing a scientific study using big data. You're looking at a lot of data about health and trying to make a determination about what causes this disease.

It's a different thing when you're testing it on people, and it's tough to tell when you actually have a false positive or a false negative. So I think from a technology perspective, we need to make sure that the underlying technology is really working as we think it should.

CHRISTOPHER OLSEN: OK. Thank you. Michael?

MICHAEL SPADEA: I think it's really too early to tell whether or not there's a gap in the regulatory regime. And even if there is one, we then have to go back to the harm discussion and define that. We really haven't agreed on what harm is. How can you have a discussion to determine whether or not there's a gap or what the remedy is if you don't know what the harm is that you're trying to protect? So I think that's one of the key places where the conversation needs to start.

We heard a lot about risks today, and I think you can always prove a point with some anecdotal stories. The goal is not to develop a perfect regulatory regime. If you went out and tried to do all the thinking to put in place a regulation that will prevent every single type of harm, that would pretty much just kill the economy. That's not the goal.

How do we allow big data and emerging technologies to deliver the greatest amount of benefit with the least harm to consumers? Obviously there ought to be a threshold of harm. Where there's a lot of benefit but a lot of harm, that's probably not a very good idea.

But even where we think that there needs to be some regulation or some remedy, just as Jeanette pointed out, we should be looking first to what are the least amount of intervention into the economy that is necessary, and then sort of gradually increase the level of intervention as necessary. I think we have a little ways to go before we have evidence that there's a regulatory gap.

CHRIS WOLF: So I think the other Chris really hit it on the head when he says that data is neither, per se, good nor bad. And I think that really ought to be the guiding light here, because we have seen that there's enormous potential for good with the use of big data. And I think we're going to get into this a little bit later, but thanks to Mark MacCarthy for previewing the study that the Anti-Defamation League and the Future Privacy Forum did on the beneficial uses of big data to identify discrimination and then, therefore, to come up with remedies for it, and also big data as a tool to fight discrimination.

And so this is the baby in the bath water theory of regulation that I typically espouse. We need to be careful when we're identifying potential problems or even real problems in regulating in a way that throws the baby out with the bathwater and that might have the unintended adverse consequence of inhibiting the positive uses of big data. And I know the FTC has that in mind. It's been, I think, a balanced day, and I'm hoping this panel will continue to be a balanced discussion of that issue.

CHRISTOPHER OLSEN: Thanks Chris. And I wanted to follow up on two different comments and one that Jeanette made about self-reg. And I just wanted to pose a question following on Latanya's presentation and ask about the Omega Psi Phi example-- the ads being shown related

to getting your arrest record, hiring a criminal lawyer, perhaps getting less advantageous credit card offers. Where does the self-reg fit in that scheme? Given your position in the industry, what's your explanation for that particular scenario, understanding you don't have any of the facts other than what was observed.

JEANETTE FITZGERALD: Exactly. I have no facts, but I could--

CHRISTOPHER OLSEN: But you know, I think you can see the website has a particular demographic, and there are particular ads being delivered. Something is going on in the machine somewhere. Where does the self-reg kick in there? Is that anecdote potentially harmful, troubling, concerning to you? Is there a role for self-reg there that would address that scenario?

JEANETTE FITZGERALD: So as you clearly stated, I don't have all the facts, and there's a lot more questions that I had just listening to the bits that they had that could, in my mind, explain some of the variations-- things like how much does the actual ad space cost, are the publishers charging different amounts for different ads? And some of those advertisers may not want to pay that different charge. Maybe they have different volumes. Whatever. There are many, many, as far as I'm concerned, factors that could be involved there.

If there was an advertiser that was not using those services for marketing services, which is what my industry does-- what my company does, we only use data for marketing purposes, period. We don't use it for any of those. And yes, we do have in our contracts, you cannot use it for any of those prohibited reasons like FCRA, and we do check and see how people actually use it.

But in my mind, self-regulation says if you're part of this industry, if you want to be part of these groups, we are going to use the data in a responsible way. We are not going to try and violate anybody's rights. But we're only going to use it for marketing purposes. Because in the end, it's an advertisement. It's the same thing you get on TV. It's an advertisement. You can either take it or leave it. If you want a different offer, go to a different bank. You don't like that bank or you want to see if the bank has something else to offer, go talk to the bank.

So there's many choices if you're marketing because all these are offers. We don't do things that are, we're going to give you credit.

CHRISTOPHER OLSEN: OK. So given that, it was simply the delivery of ads. It doesn't present an issue that the self-reg guidelines or--

JEANETTE FITZGERALD: Yup.

CHRISTOPHER OLSEN: OK. Does anyone want to address that or comment on it before we move on? Oh, silence.

CHRIS CALABRESE: I guess I'm a little skeptical. I'm not sure this is directed at self-reg, but I'm a little skeptical of the idea of "it's just marketing" actually gets you all the way to where you want to go.

One of the things in Commissioner Brill's concurrence to the recent data broker report, she talks about the use of aggregated credit scores. I'm not entirely sure I understood what that actually means, given what I understand how a credit score works. But the idea is that you're averaging credit reports in five to ten households in a specific geographic area. And I presume that you are using those for things like marketing and determining what kind of ads you are going to share with people.

To my mind, if, given the segmented and personalized nature of today's internet, if we are replicating the geographic segregation in our society and people are seeing based on what neighborhoods they are in different types of ads and offers, that is problematic, full stop. Even if they can go to another bank, if all they are seeing are the crappy credit card offers again and again, and maybe because they are the kind of people who go and get-- that's the only advertisement they see, they don't know to go to another bank. They think that's the offers they get.

So to my mind, that kind of stuff is where a market failure exists, where the CFPB should push harder to see if those offers are actually dissuading people from getting credit, or if they are ending up with worse credit offers because of them. And so that's, I think, an area to push. And I just don't think the industry self-regulatory model can fix that. Now, aggressive regulation may be able to. But I just don't think that saying it's just marketing is enough to answer those kind of criticisms.

CHRISTOPHER OLSEN: Jeanette, did you want to respond to that? You don't have to.

JEANETTE FITZGERALD: Am I allowed to?

CHRISTOPHER OLSEN: It's totally up to you. I invite you to.

JEANETTE FITZGERALD: Well, my comment to that would be, number one, not all advertising is about credit cards.

CHRIS CALABRESE: True.

JEANETTE FITZGERALD: OK? And not all advertising is determined based on aggregated credit score, which I'm not really sure I could tell you that either. I've learned about zip plus four, but I haven't learned about aggregated credit scores, and we've tried to stay out of the credit because we don't want to do any of that even with the bank and clients that we have. We're just marketing.

But it's the same theory that if I live in an apartment, somebody who's advertising lawnmowers doesn't want to waste their money and their time sending information about lawnmowers to me because I live in an apartment. It's the same sort of activity that's going on, at least from our standpoint.

CHRIS CALABRESE: But it's not. I'm sorry, it's just not. It's like, if you are looking at the credit scores of different people in the apartment and aggregating them-- which means if Dan's got

much better credit and Chris has got much better credit and I've got worse credit and I am bringing them down and they're getting worse offers, that is not the same thing. It's not the same thing as where they are or whether they can mow a lawn. It's different.

So if these practices are occurring-- and I hear a lot of, well, I haven't heard anything that's happening. This is happening. It's been demonstrated. It's up to the regulators to tell us how it's being used so that we can see if it's got this pernicious effect. I'm sorry. I interrupted you.

JEANETTE FITZGERALD: That's OK.

CHRIS CALABRESE: I apologize.

JEANETTE FITZGERALD: That's OK.

CHRISTOPHER OLSEN: Dan, did you want to jump in?

DAN CASTRO: I just wanted to say, I mean, you started the question by saying where are there market failures. Chris, I don't disagree that that could happen, but the question is, is that a market failure? Because if the three of us are living together in an apartment and we're getting--

CHRIS CALABRESE: --separate apartments.

DAN CASTRO: --separate apartments--

[LAUGHTER]

--just to be clear.

CHRIS CALABRESE: It's fine. I'm just [INAUDIBLE] to share.

DAN CASTRO: I was thinking three bedrooms.

CHRIS CALABRESE: Right.

CHRIS WOLF: Or as they would say on Seinfeld, not that there's anything wrong with that.

DAN CASTRO: So if we're sharing an apartment building--

CHRIS CALABRESE: And that was actually my larger point.

DAN CASTRO: But the point here is, though, what's going to happen over time here, right?

CHRIS CALABRESE: Right.

DAN CASTRO: Because the question is, if I'm getting worse offers or you're getting worse offers, then there's a market opportunity there. So there's an opportunity for another company to

come in and steal this business. And that's something good. That's the kind of innovation we want to see, and so that's not a market failure. That's a market opportunity.

And so if we're talking about what's going to happen over the future-- this is the panel that's looking forward-- I would say in your situation, we're going to have market opportunities where companies have the opportunity to come in with better data and solve these types of problems.

MICHAEL SPADEA: And I would say, actually, you are seeing that. You're seeing, for example, traditional large financial institutions are stepping back a little bit from low-income areas with their providing financial services. But at the same time, although it's not enough yet in my opinion, you see community banks stepping in, too, and helping trying to serve those where the large banks are pulling back.

And also you see-- again, I think back to Chris's point where you can see big data being part of the answer, you see startups coming up with and looking at alternative data points to better determine who's a good credit risk. So big data is also part of the solution, I think, to the potential problem that you're outlining. And again, I think you do see the market responding to the problem that you pointed out.

CHRISTOPHER OLSEN: Michael, let me ask you a question about something you said earlier on in your first answer. I think you had mentioned that it's premature to determine whether market failures exist where there are regulatory gaps. I think you said, and obviously you'll correct me if I'm wrong, but that more work needs to be done to define harms, to figure out what is harmful, which is a theme, I think, we've heard several times today.

And so I would just posit this question to you. If more work needs to be done to figure out what is harmful, or to put it another way, what is inappropriate, what is unethical, if more work needs to be done there, what are companies doing today in this state of uncertainty? Are they being cautious? Where are their guidelines for how to act in terms of appropriateness, ethical behavior, or fairness?

MICHAEL SPADEA: I would change the question slightly about the uncertainty part to call it-- this is a very, I think it was pointed out earlier, a very nascent industry. It's brand-spanking new, really. And I think everybody's trying to feel their way along about the risk-benefit, what's ethically appropriate. I think we need to hear more from economists as to the risk-benefit analysis.

What is the economic impact that new regulation may have? Will it promote trust unless there's benefit there? What is the economic benefit or loss to consumers who have to spend time trying to remedy inaccurate information? What's the drain there from time, money? For middle-class families, it's not as much. But when you think of low-income families that are wage earners taking a day or two off to deal with something like this, that could have a very dramatic impact on them.

I think we need to hear more from ethicists, and try to look at what can be taken in from-- institutional review boards were discussed, mentioned earlier. Is there some good practice there

that can be pulled in? How do we look at harm in those situations-- more from ethicists in general to help us figure out what's right and wrong. Should harm include things other than just economic harm as well?

I'm not necessarily advocating or arguing against any of these, but it feels to me that harm is really critical, because companies need clarity on what are the risks that they should be acting to mitigate. Without that clarity, it's hard to coalesce around a series of best practices.

CHRISTOPHER OLSEN: Yeah, and I think you teed up Chris's--

CHRIS WOLF: Well, actually, before I get to the--

MICHAEL SPADEA: I did that on purpose.

CHRIS WOLF: --FPF-ADL report, I just wanted to add to what Michael has said. Because recently, the Berkeley Information School folks asked 40 thought leaders what big data was, and there were 40 different answers. And I think the one slightly negative comment I will make about some of the discussion today is we're painting with an awfully broad brush in talking about big data and talking about harm as the same thing in all contexts. And this really builds on what Michael said.

I think we have to look at it on a case-specific basis. If there's predatory lending, predatory financial practices, that's one area to look at. If there is use of big data inappropriately to categorize people because their medical conditions, that's another area. If it's for advertising versus actual financial offerings or credit scores, those are all different things, and I think we have to consider these issues separately.

And so, to help do that, the future Privacy Forum just published something called Benefit-Risk Analysis for Big Data Projects, which tries to provide a framework that can be used across the 40 or more instances of big data and the many potential uses and harms, and really moves privacy impact assessments forward to talk about data benefit analysis. And so I commend folks here and those watching to take a look at some of the work that my colleagues, Jules Polonetsky, Omer Tene, and Joe Jerome, have done.

CHRIS CALABRESE: Can I offer a countervailing report? I've read Chris' report. I think it's very good. Everyone should also then read David Robinson's report, which I think also tackles very specific and concrete examples and, I believe, takes a little bit of a more critical view of some of the areas. I mean, everybody in the civil rights community agrees that data is a good thing and can help things, but David talks a little bit about some of the complexity of algorithms. So it's David Robinson's report. Sorry. I just thought I'd balance it.

CHRISTOPHER OLSEN: That's fine. Chris, I have one--

CHRIS WOLF: They're available as a box set.

[LAUGHTER]

CHRISTOPHER OLSEN: Chris Wolf, one follow-up question on your big data risk-benefit analysis. Are companies engaged in these sort of activities today? Do you know? Are they undertaking a sort of risk-benefit analysis today? And if not, why not? And if we think it's a good idea for them to do that, how do we go about--

CHRIS WOLF: So obviously I can't speak for all companies. I can tell you from a very unscientific sample of the clients that I advise that they are, because either based on my advice or because they came to the realization on their own, they understand that they're under the spotlight with respect to the use of data by advocates, by regulators, by the media, and, of course, by consumers.

So there is a new era of transparency that I think we can all applaud and embrace that the fact that we're here and we're talking about these things. And the fact that it is in the public policy consciences means that companies understand they have to do it correctly, that this isn't the wild, wild west, and they have to behave responsibly and do the kinds of use analyses that reflects an ethical, moral, as well as, of course, a legal judgment.

CHRISTOPHER OLSEN: OK. And how public are those analyses?

CHRIS WOLF: Well, often they're not, because often they reflect business strategies and trade secrets and products in development. And so I don't think you can expect them to be public.

CHRISTOPHER OLSEN: So how do we-- you mentioned transparency. How do we solve the transparency issue if these sorts of analyses are not transparent?

CHRIS WOLF: So this room is full of lots of different kinds of people, but among them are the corporate representatives of a lot of the folks that I'm talking about. There's a big privacy public policy community. The IP Privacy Academy that's taking place in San Jose later this week completely sold out, and believe me, there's plenty of discussion about how to do this better-- how to do the cost benefit analysis better and a lot of information sharing. And I don't think you can ask much more from companies about that.

CHRISTOPHER OLSEN: Jeanette, what are you seeing? Is that something than Epsilon does, the sort of cost-benefit analysis?

JEANETTE FITZGERALD: Yes.

CHRISTOPHER OLSEN: And just to key off my earlier question, I think there was discussion in one of the earlier panels, I can't remember which one, but I think it was Danah Boyd who said there's a lot of public uses of data and data sets that are very transparent. How the data is crunched, how the data is analyzed, what the results are, all of that is made public.

There is no similar transparency on the commercial side. I think the cost-benefit analysis, the benefit-risk analysis sounds like something responsible companies should be doing today. The question is, how does anyone get any sort of comfort that the analysis is not affected by-- either

not affected by concerns we would care about or the results aren't unfairly impacting someone? How do we get over that transparency hurdle?

JEANETTE FITZGERALD: So there are a couple things. One is certainly how we look at any new products or how we continue to use products has evolved over time, because this notion of privacy and how society accepts it has evolved over time. The "privacy profession" hasn't been around that long when you look in the scheme of things of how long businesses have been operating. Nobody really thought about it that much.

So it's evolving now. And as Chris Wolf said-- there's too many Chrises here-- I mean, IAPP, it sells out all the time. There's always people that are-- I see the same group all the time, but we're all talking about new issues as it evolves.

If we as a company, my company, for example, Epsilon, decided to tell everybody exactly how we did a risk-benefit analysis, that would be giving up trade secrets. We're not going to do that. Other companies are going to feel the same way. That's part of our quote unquote "special sauce" to make it.

Now, that doesn't mean that if there was an impact that somebody felt was discriminatory, that somebody is not going to come back to us and say there's a problem here. And then what happens? Our name shows up somewhere. We don't want our name to show up.

So there's a lot of good reasons why we're very careful about those things. Our team looks at things like, if you were a consumer and you had given your data in this first instance, for whatever reason, would it make sense to them that they would be using it this way later? Now some of it's complementary, and you can figure out, OK, it makes sense. Some of it is so far in left field you just have to look at your team and say, I get what you're trying to accomplish, but this ain't gonna work. Not going to do it, because we can't explain it later and do it with a straight face.

I mean, the fact that there are hearings, the fact there's a huge group of people in the public and a lot of them are sitting around here who will come and look and tell you're doing something wrong is pretty good at making sure that you do the right thing.

CHRISTOPHER OLSEN: Does anyone else want to offer anything on the transparency concept? How do we improve the state of transparency of data use or analytical tools or algorithms today?

CHRIS CALABRESE: I would like to just-- I think we're sort of woefully inadequate when it comes to transparency right now. And so I'll pick on the data brokers just because there was a recent report, so there's a fair amount we know about exactly what their practices are. But in the recent FTC report, I believe it was Axiom, they said Axiom had something like 1,500, or more than that, data points on every consumer.

I went and looked at my Axiom profile. There's nowhere near 1,500 data points about me, nor, and I think more importantly, is there anything about how they're being used, like what score or

assessment am I being offered? Am I being grouped as an urban scrambler? Am I a vulnerable consumer? In my personal life, I'm very vulnerable.

But until I know those assessments, and I get that it's special sauce, but the individual consumer should be able to know if they are being targeted or there's an assumption being made about their health or their finance. So that's the kind of transparency I'd like to see more of.

JEANETTE FITZGERALD: So let me throw one thing out in response to what you're saying that Axiom does. Epsilon, too, has a section within our site, I don't know what it's called, a microsite or not, because I never am sure what those are, but where a consumer can go and we give them education about how their data is used for marketing purposes, places they can go with the DMA, places they can go to the FTC to get further information. They can go to a couple different other sites that will give them lots of information about how it's used.

I don't know how many people are actually going to go read all that stuff. It takes a long time to slog through it because we all did before we put it up there. But the other thing it does is it says, look, if you want some information about you, if you want to know what kind of information we have about you and what group we put you in, you have to send us something because at this moment, we haven't figured out the right way to do it online. I've got a couple ideas, but I don't like what I've seen so far. Because we don't have credit cards, we don't have driver's license, we don't have social security, and I'm not going to ask for those to then give you a report. That seems counterproductive.

But we can give you a report, and I should have brought it, and I'm sorry I didn't. But it basically says your name, your address, whether you have kids in your house, whether you own your home, and then some of your interests. So we like to be outdoors, yes, we do hunt because we're Texas, fish. And then it says at the bottom, you're in this group. And it says these are the basic characteristics of this group. The salary's about this. You buy books and magazines. You shop online. You-- I can't remember what the other ones are.

But that's available. You can go get that and we'll show you. And frankly, after you read that, if you're really still worried, we'll opt you out.

CHRIS CALABRESE: And I appreciate it. And I've seen some of it. I've gone to-- I don't think I've ever been to yours, but I also don't see the same alignment when I read about what regulators are writing about this industry. And that makes me wonder where the disconnect is between-- and maybe it's a classic, we've got good actors and bad actors. Somebody brought that up too, which is of course, the classic argument for regulation, is that the good actors are already behaving properly and the bad actors aren't going to do anything unless you regulate them. But I do see a disconnect in the transparency.

MICHAEL SPADEA: If I could just briefly-- I don't know if the answer is more transparency, but perhaps better transparency. We've all heard about the studies where it would take you 29 years or something to read all the privacy policies. It's not that consumers can't do it if they want to, it's just who has the time to do that? If we inundate consumers with descriptions of the technologies and the business process and all the data flows, they're not going to read past the

second privacy policy. Just think about when you buy a new computer and reinstall your software and re-update the stuff from the cloud. You're not reading all that. I don't read all of that.

More transparency is just going to dull the senses, which I think is what you're seeing a little bit with the breach notification piece. If you think about the airline industry, you don't go on to the website and it doesn't show you how many hours the pilot slept or what the maintenance records are for the airplane. No. Those are not the factors that you're looking at when you make the purchase. You want to deliver the critical information at the moment in time, and that's that.

Perhaps a better approach may be, which we already do in some ways, risk rate the data. For that data that's most sensitive, that might have the most potential impact, there's a higher notice requirement there. But just blanket across the board dump tons of more information on consumers, I can't see that as protecting consumers. And in fact, it may put them at more risk.

CHRISTOPHER OLSEN: Michael, and I don't disagree with that. I think in terms of transparency, there are a variety of ways to deliver that. And I think what I was contemplating is some mechanism for companies, and I think Peter Swire alluded to this in the last panel, was if there's an unfair practice, if there's unfair marketing going on, you could foresee a scenario where the business has a justification for why it engaged in the particular marketing campaign. It's not necessarily a justification for consumers. It could be a justification for a self-regulatory governing body like DMA. It could be a justification for regulators.

It's not necessarily giving a notice to consumers every time they receive an ad that says, this ad was delivered based on the following 15,000 analytical data points. I don't see that as being particularly helpful.

But I think if there is a concern about how data is being sliced and diced and crunched, and whether there's something going into the analysis that is of concern, or something coming out at the end that raises questions about the transparency of the analysis itself.

So Jeremy, and Dan too, but Jeremy, I'd like to ask, is there a role for technology, for example, in helping address some of the transparency issues or some of the concerns about whether there's something of concern happening behind the curtain?

JEREMY GILLULA: I mean, I certainly think so. I think the technology can help a lot. I mean, going to what you said about you don't necessarily need to show every consumer exactly how they got this ad unless maybe one or two consumers, or some consumers who are concerned are interested. And then if there were a way for them to click on a little part of the ad that said, hey, yes, this is why we served it to you.

So it's not that everyone always gets it all the time, but so that people who are concerned can try and understand. If there's someone like Dr. Sweeney can, when she's doing her investigations, not just say, OK, this is what we saw, but, hey, and this is why the ad companies say they gave it to us. And I think through a little bit of disclosure through the-- it's not a technologically infeasible thing to try and do.

In terms of also just using technology to determine when discriminatory things are happening, it also occurs to me that in some way-- and I don't know if this is the sort of thing the EFF would take on-- but people could turn big data back on the data brokers. You could think of a browser plug-in that collects the ads that you're seeing. And then if a lot of people install that, then you can start comparing what ads different people are seeing.

And so in some way, you could essentially collect big data on big data, and then try and do some open source analysis perhaps. And I think the reason that something like that might be valuable is because a lot of the times, these sort of effects aren't necessarily obvious. Because most the time, I don't know what ads Michael's seeing. I don't know what ads Jeanette is seeing. I just know what I see, and I just assume it's normal. I assume there's nothing discriminatory going on with it. And until people can start to compare these things, I think a lot of this will sort of be shadowy and not very transparent.

CHRISTOPHER OLSEN: So I want to come back to the technology in a second, but I want to let Dan jump in here.

DAN CASTRO: Yeah, actually building off of what Jeremy said, I do think big data is the solution here to many of these types of transparency things. In fact, what you're describing, for example, is the shoe car model, where it's a company that collects all the data from car dealerships about what prices people pay, and if you want to use them, then you get to find out what other people paid. You share your data, and that's maybe a less discriminatory way of buying a car. You know you're not getting sold based on the type of shoes you're wearing when you go to the dealer.

So there are lots of ways that you can use this. And this gets to a really important point about whether the discrimination or harm that you're positing here is something that's intentional or unintentional, which is something that the first panel talked a lot about.

And how you address those might be different. So you need to think about which type of problem you're trying to address. If it's unintentional harm, you really do have to address a lot of that through data analysis. Nobody's intending to do this, so you have to make sure you have smart data scientists doing things consciously, but also that you're able to evaluate outcomes. If it's intentional, then you have a human problem, and maybe you address that differently.

But the second point here is really about whether or not you want to open up these algorithms. Because I think that's really important when you start thinking about the trajectory of how innovation will occur in this space. Ultimately, I think the goal should be that you're innovating around accurate data, so that the innovation is really in the algorithms.

If you look at the Open Data movement, that's what this is all about. It's not about who has the data, which is the kind of world we live in right now. That's why we have data brokers, because you can buy data. It's really valuable, maybe, what data you pay for.

What you really want to get to, I think, is where getting access to the data is really easy, and it's all the intelligence and innovation that you built on top of that that's hard. So you want to

promote that. And so I think part of that is by allowing trade secrets to exist, is by allowing intellectual property to be protected here.

So as we think about regulation to address this issue, I think we have to consider what this data science space will look like in the future. And part of that will be accurate data. So the question then will be, do you want innovation and algorithms or not?

CHRIS CALABRESE: Sorry. I was just going to say this is hard. Take the E-Verify example. E-Verify is a government system for essentially deciding whether you are going to get a job. And the goal is to say if you are lawfully in the United States, you're work eligible. You query this government database. And if you're not, the employer isn't supposed to hire you.

Well, this database has been in existence and being perfected since 1996, so a very long time. It uses fairly homogeneous data. It's all government data. It's all relatively discrete set of data sets. It still has an error rate that is 20% to 30% higher for certain classes of people who are legal immigrants-- who are in this country legally but are immigrants.

So this is a system that's run by the government with oversight that still has substantial error. So I didn't mean to make too much of a point, but this is really difficult to do. And I think that we need to acknowledge maybe that some of our uses, especially if they're going to result in things like people not getting jobs, until we have a high degree of confidence that we're actually doing this right, maybe we shouldn't be allowed to do it. I know that's a little bit anathema, but--

DAN CASTRO: I think that's the throwing the baby out with the bath water problem that I was talking about, because the point of the FPF-ADL report-- and for those who don't know, the Anti-Defamation League was founded 100 years ago to combat anti-semitism and promote justice and fair treatment for all the future. Privacy Forum is a privacy think tank. We came together like Reese's Peanut Butter Cup to create something really good together by combining both our missions.

And so we looked at things, for example, like the Urban Institute, which recently combined public school data with demographic information to show segregation in public schools, a use of big data to identify a problem with discrimination. In another example, the National School Board Association supplemented the Department of Education research with raw census data to also show disparities in the effect of school disciplinary practices on the graduation rates of various minorities.

We've seen big data used to fight discrimination in the workplace. Somebody mentioned earlier this company Intello, which produces a digital recruiting tool for those companies who want to have a more diverse workplace. It helps them use big data to identify prospects.

In another example, Google has used big data to identify problems in its own hiring process. And it's a real credit to the company because it admitted that its own brain teaser interviews apparently unfairly favored males. And so it's now reformed its hiring practice after making that realization to evaluate candidates without questions that may unfairly disadvantage one gender.

We see the EEOC using something called FedSEP, which is an electronic portal through which more than 325 agencies now report workplace discrimination charges, and those numbers are crunched in various ways. So we have 14 examples in our report of how big data--

CHRISTOPHER OLSEN: Right. And those are great examples, and I think they all serve to really demonstrate that big data has tremendous societal value. And I think I want to-- well, let me first jump in here for a second. I think what we're talking about-- I think, Chris, you and maybe others on the panel are teeing up the scenario of regulation versus no regulation, law versus no law. And that's certainly an option. Should we have another law? Should we recommend legislation? Should Congress pass legislation? That's an option. We all know how challenging it is for something to come out of Congress.

So let's talk about best practices. Are there practices that companies can engage in that would measure potentially or cabin or restrict or evaluate potentially harmful uses that they're not going to impact the potentially beneficial big data uses? I mean, if you have a data ethicist or a chief fairness officer, that person in the evaluation that that person undertakes before a new program is rolled out, it's not going to curtail the benefits of big data.

CHRIS WOLF: No question about it. I love this quote from a report the KPMG recently did. They said, "Organizations that attempt to implement big data initiatives without a strong governance regime in place risk placing themselves in ethical dilemmas without set processes or guidelines to follow. Therefore, a strong ethical code along with process-training people in metrics is imperative to govern what organizations can do within a big data program."

CHRISTOPHER OLSEN: OK. How do we come up with a strong ethical plan?

CHRIS WOLF: I've lots of booklets to wave around, but I come back to the original one, which is the benefit-risk analysis for data projects. Having a framework and a methodology and a discipline within an organization is absolutely essential. Now, is it like a traditional IRB? Maybe not, but Professor Ryan Calo has said that it's certainly something to think about as a way to, if there is a gap, fill the gap, but even if there's not a gap, to avoid adverse consequences.

CHRISTOPHER OLSEN: Jeremy, you wanted to jump in.

JEREMY GILLULA: Before I was actually going to sort of address a question to Christopher. And it sort of goes back to what you were saying, that we're painting big data with a very broad brush. And it seems to me, the difference between what you're talking about and what Christopher Calabrese was talking about is he's talking about uses of big data where a decision is made about an individual. Every single positive use of big data I've heard so far today is, we discovered something about a population. It's not, we decided to classify someone.

CHRIS WOLF: But that then benefited individuals to allow them to have an education free of discrimination or health care free of discrimination.

JEREMY GILLULA: Sure. Absolutely. But that's called science. Just that we have more data, we collect more data, we can find more things. But what we're talking about, a lot of the harms

are harms where data about a person is then used to make a decision about that person, not, we found some trend and then we adjusted our methods. We found some trend, and then Google decided, yes, we need to change our hiring practices. It's, Google looked at your data when deciding whether or not to bring you into an interview, and based on the data, decided not to bring you in.

MICHAEL SPADEA: I disagree with that characterization. He gave examples where people benefited in the end. And your examples or Chris' examples, he was talking about harm to individuals. It's called the risk-benefit. We look at what the potential risks are, the potential harm, and we weigh it against the benefit. And you can't answer the question you're posing without the bringing of the two together. So I would say no. They are apples to apples, not apples to oranges.

CHRIS CALABRESE: Yeah, but it's easy when the benefits are to the company and the harms are to the person. It's like, yeah, great. I don't want to be the guy on the harm side. That's why we need government standing here saying, that's not OK. This data isn't accurate enough. This is harming people. You didn't give this person a job because--

CHRIS WOLF: But Chris, you're assuming that it's uniformly harmful to the consumer, and no one has said that. No one has said that today.

CHRIS CALABRESE: You need a framework in place backed by something more substantial than self-regulation to make sure that the harms are as mitigated as possible and do not fall on particular classes of people or individuals.

MICHAEL SPADEA: You're assuming that if a benefit is provided by a private company that there's something wrong with that, and that doesn't equal to an actual benefit. That's what, at least, it sounds like I'm hearing. Let me just finish one sentence.

CHRIS CALABRESE: Sure. Sorry.

MICHAEL SPADEA: I was going to add, though, I do agree to your point. Where there is harm, there should be some type of remedy. We shouldn't just leave consumers floundering. The question was, do we need a law or not? And I think what I'm trying to say is that the evidence to say that we need legislation now is not there.

It may. As this industry continues to develop and we have more information about harms, about benefits, we need to continue having this discussion. And there may come a point where we do need further regulation or legislation.

But we need more information. We should start with the least interventionist approach. And if that doesn't work, we ratchet up the intervention.

CHRISTOPHER OLSEN: And I guess the question I would ask before I turn it over to Dan, it goes back to the transparency question, which is if there is no transparency about how the data is being used, then how do we get to the, "we have more information" point in order to make a

decision? And it may be that companies internally know how they're using the data, but they're the only ones who know that. So Dan, you wanted to--

DAN CASTRO: I think this would be a good point to talk about this paper that we released last week called--

CHRIS CALABRESE: Everyone else.

DAN CASTRO: Everyone else.

CHRISTOPHER OLSEN: Let's bring the level down a little.

DAN CASTRO: --since you brought it up.

MICHAEL SPADEA: It's the afternoon panel. We did need to--

DAN CASTRO: We released a paper called "The Rise of Data Poverty in America." And this gets to what Chris was talking about, which is what are the individual benefits and risks? And so the point of this paper, we went through and we talked about specific benefits that individuals are seeing in areas like health care, education, and financial services. And we also talked about the challenges that we've had in the past, both in terms of the digital divide and how that might translate into a data divide, as well as challenges that we've had in small data sets.

So the best example of this is in health care, where historically, if you look at, for example, clinical trials, minorities and women have been under-represented in this data. And just as when we're talking about big data, decisions are made from big data, decisions are made from small data as well. And so decisions have been made in the past, for example, by the FDA about what drugs and treatments were safe and effective. And it turned out that, of course, once they released it to the full population and that population didn't match up with the clinical trial population, some things were unsafe.

And so the questions we asked in this report were what challenges might certain disadvantaged communities see if there are, in fact, data gaps-- if there are data-rich communities and data-poor communities. And we actually looked at Wikipedia contributions on a per capita basis, and we tried to do an initial mapping of what data deserts might look like in the United States. And there are these really interesting gaps.

And so the questions are, as we're using all of this data, and I think that's generally good, are there populations that are left out, and what do you do? And so if you compare what we've done with the digital divide, we don't say, oh, some people don't have access to computers. Let's stop using computers. But that's kind of the message that I hear sometimes today on the panel, which is that you don't want to use data. That's not the answer.

The policy answer to that type of problem, of the data divide, is to say how do we make sure that disadvantaged populations also have data available about them that they can share in these

benefits? Because when you look at it, it's very clear that there are huge economic, educational, health benefits. And we want to make sure all of these groups can share in that.

CHRISTOPHER OLSEN: Yeah. That raises a good point. It reminds me of the street bump example where data was being collected about road conditions from smartphones. And there was a question about how broadly representative different communities were in that data collection.

I could see something similar happening with wearables. If policy decisions on health data are made based on input from wearable devices, are there certain communities of people that are being excluded, which, again, sort of suggests that at some level there's some sort of fairness or ethical approach that has to be applied as a frame for any of these data collection practices.

DAN CASTRO: Brief response to that. I think part of that is data literacy, not only among the data scientists so they understand what exactly it is they're doing but also policymakers who are interpreting this or interpreting the results. And hopefully we're doing some of that today.

MICHAEL SPADEA: In the street bump example, it's a great example of how you can get tripped up. But it's also a good example, I think, of how you can fix the situation. So the answer would be not to get rid of the app or anything like that, but if you understand where the smartphone saturation is and where it isn't, you can then put in mitigating controls.

So we know that in areas which will be predominantly middle class or upper class is going to be high smartphone saturation, and therefore Department of Public Works is going to get really good data on where are the potholes are, and they're going to get fixed. But that's not going to happen in the lower income neighborhoods.

So what do you do? Well, you know you need to have something else in place for those neighborhoods. So you take the money that you save from pulling DPW people on pothole patrol or whatever they do, you take some of them-- you take half of them and you just take that money savings and you put it someplace else. You take the other half and just throw them right into the neighborhood that doesn't have that saturation.

At the end of the day, you get to the same place. You get there more cheaply. Everybody's happy. So you can, where you know whether a problem like that exists, and the key thing is the governance that Chris talked about earlier, that there should be a process to spot those risks. The ethics piece comes in where, OK, so we now need to fix it. We can't just let that harm sit out there. But we can still roll forward with the application, and with a private company, generate revenue and service the consumer.

CHRISTOPHER OLSEN: Let me key off that and tee up a question that we've been sort of hinting at during this panel. There's a debate between a new space model of data handling and a data minimization approach. And I'll just pose this question. There's been talk on various panels today about data governance. Chris Wolf mentioned it. You need to apply a data governance methodology. We've talked about making sure we have a clear idea of fairness. We've talked about having an ethical approach. We've talked about how we're at the early stages of these sorts of practices.

So I put the question to the panel. If we haven't resolved the framework for applying an ethical construct to data practices or fleshing out harm the way we need to, why isn't data minimization still an important component of information handling practices?

CHRIS WOLF: So maybe I can start. And I referenced the fact there are at least 40 different definitions of big data. But one kind of fundamental understanding is that it relies on volume, variety, and velocity of data that leads to unexpected discoveries. And so how do you provide notice at the time of collection and allow consumers to make choices about discoveries that you don't know will happen? That's conceptually one of the problems I have with this idea of a collection limitation.

But I think a more practical issue is one that I think Pam Dixon very candidly acknowledged, is that there are huge data sets already out there, structured and unstructured-- data exhaust as she referred to it-- and even if we are able to minimize data collection or to provide options that put limits on the collection, we're still dealing with huge issues of use. And, as we've discussed here today on this panel and others, consumers aren't simply going to take advantage of the transparency options and make the choices that perhaps we think they ought to. And there has to be someone responsible in the ecosystem.

And that's why we urge the governance model and the focus on use without rejecting the FIPS of collection, but without unduly placing emphasis on it.

CHRISTOPHER OLSEN: Anyone else want to comment on this?

CHRIS CALABRESE: Yeah. You'll be shocked to learn that I think that use is not enough in and of itself. I also think-- and I think that data minimization has an important role, but I guess I would put a plug in for all the FIPS. I mean, the fact is that a lot of times, consumers don't take the time to know about what's being collected about them because there's nothing in it for them. All they can do is learn about it and they go, ah, well, you're out of luck. You don't know their rights to do anything with this information or limit it.

So I think that having both minimization but also use limitations and the ability to, for example, say I'm going to keep my salary information from becoming part of this data ecosystem because I am noticing that I'm not getting as good of coupons and offers, because people think I don't make enough money to be worth those good offers. I want to keep that information to myself.

Now, if you have the ability to control various types of information, I think you are much more likely to learn how it's used, and much more likely to endeavor to be an active data user, at least about yourself.

CHRISTOPHER OLSEN: Anyone else on this particular point?

DAN CASTRO: Well, just to a couple things that have been mentioned here. I think we have to separate between harms to an individual as in, I'm paying more than I'm paying today versus I'm paying more than I'm paying today because someone has something wrong about me. Like the E-Verify example, if I can't work because the government fundamentally has something wrong

about me, regardless of the law itself, but that's a different problem than if my insurance company charges more because I speed a lot and now they know about it.

We need to separate out those types of-- one is a harm. One is actually an improvement, because there's actually a benefit to somebody else.

CHRIS CALABRESE: There's an information asymmetry here, right? If I know that you are wealthy, and you are more likely to come into my store if I give you really robust coupons, say a \$15 coupon, but if I know your income, and I know that I don't have to give you a \$15 coupon. I can get you in the store with a \$3 coupon. Now we can argue about whether that's a genuine harm or not, and you can shop somewhere else. But the fact is that you know something about me, and you're using that to provide a differential in something that I would value, which is to say--

DAN CASTRO: But the point of that is, though, you can do the opposite as the consumer. So a great example of this is if you look at car dealerships. And it used to be if you wanted to get a used car and you didn't have many assets and you didn't have any collateral, you weren't going to get a car. They weren't going to make a loan to you. And the reason is because you would have a car, and you could drive off with it and stop making payments. And there was this huge risk. No one was going to do that.

Now, using data, you can actually say, OK, I'll have a GPS-enabled device. I will tell you where I am. You can have this data about me, so I will prove that I'm not running off with the car. That way, there's a significantly lower risk to you, and then now you'll make a loan.

And so you have all these dealerships that are now making loans to individuals where they didn't have access before. So if you're a single dad, you get a job, you need reliable transportation, now you can do that. That's the consumer using data for good. And that's what we want to see more of.

CHRIS CALABRESE: But that's a voluntary-- I'm choosing to give you that data in response to a need. That's completely different than my unwilling disclosure of my salary through a third-party data broker. It's apples to oranges.

CHRISTOPHER OLSEN: Let me turn to one point we've touched on a bit earlier today and tee it up this way. Data governance, it seems like everyone agrees, is important. Companies are moving towards more formal risk-benefit assessments, which seems like a good step. But we've discussed at length the transparency issue there. It may not be visible how companies are applying the data governance methodology.

And so should consumers consider other options that exist, or should we consider other options from a technological standpoint? Should we push for data tagging, for example, that would identify the provenance of data elements, or are we beyond that? Or should we consider random identifiers that would mask your identity as you navigate the web so that you appear to be a new person every time you visit a particular website? Or should we entrust our data to a third party with a permissions scheme?

Are there measures consumers can take or companies can deliver that would mitigate the risk that the data would be used in harmful ways?

CHRIS WOLF: Unlike your first question, which I refused to answer as a yes or no question, the answer to all of your questions is yes. I think technology does have potentially a very significant role here to play to provide exactly those kinds of protections, exactly those kinds of options. You didn't say de-identification specifically, but I think that was implicit in your question about random identifiers.

So I think there's great hope in technologists. They've certainly gotten a lot richer than lawyers. So I think there is--

CHRISTOPHER OLSEN: Even you, Chris?

CHRIS WOLF: That's private information.

DAN CASTRO: Axiom knows.

CHRISTOPHER OLSEN: Anyone else want to address the-- Michael?

MICHAEL SPADEA: I think it ties in nicely to the responsible use viewpoint, in that if you own it, if you have the data, you're responsible for it. And I would interpret that quite broadly. You're responsible for who it's transferred to. You're responsible to keep it secure. You have to act in a responsible manner. And implicit in all the risk mitigants that you just set out there, those would all be tools in the toolbox of the organization to help mitigate these risks. They need to act in a responsible manner.

And I would just add, I think, actually, the responsible use viewpoint requires a strong and well-resourced regulator, because they're the ones at the end of the day that are going to really have to make some of the determinations about what's responsible. I don't know if this is true or not, but somebody from the FTC was telling me that the FTC is resourced at, today, 50% less than it was in the 1970s. And if that's a true statement, I'm shocked.

So I would say I really like the reasonable use. I think it ties in exactly-- or responsible use-- to what you've just said. But the FTC needs a little more muscle to make sure data is used responsibly.

CHRISTOPHER OLSEN: Anyone hear that who has the purse strings?

CHRIS WOLF: Michael will be taking up a collection at the door.

CHRISTOPHER OLSEN: We have, I think, just under five minutes, and I'd like to ask each of the panelists to, in their closing, make recommendations to anyone they want. You can make a recommendation to industry, to the FTC or other regulators, to Congress, or to consumers. What would you recommend are the next best steps to take as we move into the world of increasingly complex algorithmic analysis? I'll start here, and we'll move down.

CHRIS CALABRESE: My recommendation would be that regulators, specifically the FTC but especially the CFPB, very aggressively investigate whether the Equal Credit Opportunity Act does reach some of these practices, especially the marketing practices and the marketing of credit offers, and whether the marketing of higher credit offers to particular segments of the population, in fact, discourages those populations from pursuing credit offers, and hence, violates the Equal Credit Opportunity Act.

And I will do my own little plug and say I think that the ACLU will provide more formal written comment on this and encourage this before the close of the comment period.

CHRISTOPHER OLSEN: Thank you. Dan?

DAN CASTRO: So I'd say I think this is definitely the start of the conversation. We need many more voices here. I think it's interesting. Today's workshop has been fantastic. But across town, there's a predictive analytics government conference going on with some of the best predictive analytics data scientists in the country, and they're not in the room. And so we need them here. They certainly should be part of conversation. And I guess my recommendation here is that we really need to be thinking about the benefits here.

To me, if you care about discrimination, if you're worried about health care or improving education for our kids, the biggest risk is not how data is being used. It's that we won't use it enough. And we need to figure out a regulatory environment and policy recommendations to help encourage more use of data.

CHRISTOPHER OLSEN: Jeanette?

JEANETTE FITZGERALD: So I would say that we need to spend time figuring out the best way to educate consumers about the data that's being used on them, about them. And it's not just how it's being used, but we also need to teach them that they can talk to the companies that have their data. They can ask questions. And those people will help them understand what information they have and how it's being used. And I would encourage any other company that's been thrown in the data broker realm that they, too, think about ways that they can show consumers the information they hold on them and how it's being used-- what category people fall in.

JEREMY GILLULA: So I would build off what Jeannette said, and that I do think that getting consumers educated about these sorts of things would be a huge benefit. I think part of that goes to the transparency we've been talking about. And I think it would actually be a benefit for data brokers and marketers to be a little more forthcoming about that sort of thing.

Because right now, if you try to go and find this stuff, it feels like diving into a deep and shadowy world. And that may not be what they mean it to be, but that's what it feels like. And so being more transparent about the-- and I realize a lot of this is trade secret, secret sauce, but even just sort of giving general ideas to consumers, I think, would be a huge benefit.

The other recommendation that I would make is actually towards the FTC and, really, anyone sort of observing this space, and that is, look closer. Look past the hype. I'm going to reiterate

this point that I said earlier because I don't think it was adequately addressed, that a lot of the benefits we're seeing that people tout about big data are benefits that come from analyzing and learning things about a population. And for every 10 benefits of big data I hear about that, I hear maybe one about how individualized targeting via big data helped people.

And it's that individualized targeting where I think a lot of the harm is. I don't think there's a lot of harm in learning about, hey, look, these types of students need help, or these interviews or are harming people. It's when decisions affect individual people's lives that I think we need to start thinking about.

CHRISTOPHER OLSEN: Michael?

MICHAEL SPADEA: I would urge companies to develop enterprise-wide risk programs, and as part of that, have a data risk framework. I think you could just simply read all the papers that have been discussed or otherwise provided today to come up with a list of the potential risks. And you make determinations about what apply to your organization. Make determinations about where your risk appetite is, and put controls in place.

While I get some of the questions are difficult, a lot of it's not rocket science. Come up with the risks. Put the controls in place. Test against them. Have good governance in place.

I would say to everybody we need to have a discussion about harm. I think that's central to how we move on from here.

To the FTC, and perhaps all regulators that play in this space, these workshops are great. It's been mentioned that we need to bring in some more of the economists and the data ethicists and the scientists, and so on. And so I wonder-- everything just moves so quickly. I mean, it's like we should schedule the next big data workshop a year from now and just schedule it now and get it done.

Maybe we should be having an information week where we're talking about best practices and privacy one day, security the next, data governance in general, a piece about educating the FTC and everybody else about the technology and the business models. It's kind of like Shark Week, but if you could combine them, we'd have consumers tune in and there'd be the education piece right there.

CHRIS WOLF: Is that another dig at lawyers?

MICHAEL SPADEA: No.

CHRIS WOLF: Just wanted to make sure.

MICHAEL SPADEA: My lawyers are my best friends.

CHRIS WOLF: So I'm hoping that just as privacy by design is entered into the lexicon of all privacy professionals that data benefit analysis or benefit-risk analysis with respect to the use of

big data will also become something that's reflexive and something that every privacy professional talks about. And I think that will avoid a problem I see with Jeremy's analysis of focusing on who benefits. Because if you put rigid, one-size-fits-all restrictions on the collection and use of data, you're not going to have benefits for anybody.

CHRISTOPHER OLSEN: Well with that, I would invite the audience to thank our panelists for a lively discussion.

[APPLAUSE]

Thank you guys. And Jessica Rich, Director of the Bureau Consumer Protection is going to give closing remarks.

[APPLAUSE]

JESSICA RICH: I'm going to sit in one of Christopher's spots, because this podium is too tall for me. So good afternoon. Many of you are still here, I see. It's great. We've had a really great day of discussion and debate regarding consumer protection issues surrounding big data and, in particular, its potential impact in certain consumer groups.

My remarks will be short and sweet. They're never quite as short as I think, but they'll be short and sweet because I know it's been a long day for everybody.

First, I want to thank the team, many of whom are sitting over there, that put together this terrific event.

[APPLAUSE]

Tiffany George, Katherine Armstrong, and Chris Olsen here from the division of Privacy and Identity Protection. Katie Worthman, Patrick Eagan-Van Meter, and Malini Mithal from our Division of Financial Practices, and Jessica Skretch and Lesley Fair from our Division of Consumer and Business Education.

And also thanks to our event planning and web teams, our press office, and honors paralegals for all of their help. It takes a lot of people to put these on. And thanks, of course, to our great panelists and our audience and all of the folks who we spoke to and met with as we were planning this event.

So this workshop was part of the FTC's ongoing program to examine emerging or growing consumer protection issues. It was an inevitable follow-up to what we learned at our seminars on big data last spring, what came out of our data broker report, and what we learn every day by just opening up the paper-- yes, I still get a paper delivered to my door, a hard copy-- and following industry developments.

Today we learned about many beneficial uses of big data. For example, we heard case studies about how big data can help fight discrimination, predict the risk of homelessness, increase

diversity in the workplace, help ensure certain populations are getting the health care they need, and actually empower traditionally vulnerable populations.

But we also discussed the risks that big data can lead to selective opportunities, stigmatization, and discrimination. For example, Latanya Sweeney presented some interesting preliminary questions about how big data may impact the ads that visitors to certain websites see based on the presumed race of the visitor. Solon Barocas discussed the ways in which existing patterns of discrimination inherent biases present in the use of little data, such as the categorization of consumers based on their likelihood to buy can be replicated with potentially greater scope or scale in the use of bigger data.

Other panelists talked about how predictions developed for one purpose, such as whether a person will drop out of school or buy a particular product, could be reused for more harmful purposes, or as a proxy for income level, race, or other characteristics.

We discussed many important questions for which we need to continue seeking answers. How will big data be used for marketing, fraud detection, or the eligibility for various offers? How do existing laws apply to big data? Even apart from laws, how do traditional approaches to privacy apply to big data? Are transparency and choice still important and feasible in this environment, and what about data minimization and data [INAUDIBLE]?

We also discussed what happens when certain populations don't have the same sort of access to technology as other consumers. Will inequalities result from this lack of collection and use of data that could otherwise provide benefits to these populations? And we began, but hardly finished, discussing the overarching question that was the basis for this workshop, how will all of these new and evolving practices impact certain populations, and what steps can and should businesses take to make sure particular groups are not disproportionately or negatively affected?

I think it's fair to say everyone here today agrees that big data is not going away, and it's only going to get bigger. Our collective challenge is to make sure that technology continues to provide its many benefits and opportunities to consumers while adhering to core consumer protection values and principles.

To that end, our chairwoman this morning outlined three steps for moving forward, which I'll emphasize as my parting message. And actually, these three steps, or themes, or whatever you want to call them, were echoed down the line in this last panel. First, as a law enforcement agency, the FTC will work to identify areas where big data practices violate the laws currently on the books that we enforce, including the FTC Act, the Fair Credit Reporting Act, and the Equal Credit Opportunity Act, and will bring enforcement actions where appropriate.

Second, we will continue our efforts to examine and raise awareness about the consumer protection concerns surrounding big data through speeches, consumer and business education, which we certainly need to do more of, and potentially, follow-up events or a report on this workshop. And third, we will encourage businesses to design their analytical FIPS systems with an eye to the concerns we've discussed here, avoiding bias or disparate adverse impact on particular populations of consumers.

Finally, I do want to mention that our comment period will be open until October 15. Please don't be shy. Please comment if you have something to say. You can file comments electronically or by paper. The details are on our website. And with that, let me just thank everyone for coming, and have a great evening. Thank you.

[APPLAUSE]

[MUSIC PLAYING]