FEDERAL TRADE COMMISSION

SPRING PRIVACY SERIES

ALTERNATIVE SCORING PRODUCTS

MARCH 19, 2014

Federal Trade Commission
601 New Jersey Avenue, N.W., Conference Center
Washington, DC

Reported By: Stephanie Gilley

FEDERAL TRADE COMMISSION

I N D E X

W E L C O M E

MS. ARIAS: Hi, everyone, and welcome. Thank you so much for joining us to the second installment in our spring privacy series.

Today we are going to be talking about alternative scoring products, but before we begin, we have some small administrative and security issues that we have to cover before we begin.

All right. So please note that anyone who goes outside today without an FTC badge will need to go through the magnetometer and the x-ray machine again prior to reentry into the conference center.

In the event of a fire or evacuation of the building, please leave the building in an orderly fashion. Once outside the building, you need to orient yourself to New Jersey Avenue. So across from the FTC is Georgetown Law Center, look to the right front sidewalk. You need to check in with the person accounting for everyone in the conference rooms there.

In the event that it's safer to remain inside, you will be told to go where inside the building, so just listen for an announcement. If you spot suspicious activity, please alert security. Security is right outside, where you came in.

This event may be photographed, videotaped, webcast, or otherwise recorded, so by participating in this event, you are agreeing that your image and anything you say or submit may be posted definitely at FTC.gov or one of the Commission's publicly available social media sites.

All right. So to submit questions today while the event is happening, question cards are outside, available right on the table where you came in. So if you want, just go ahead and grab one and write your question and then we will have one of our fabulous paralegals here at the FTC walk around and collect them and bring them up to us and then we'll ask the questions to our fabulous panel coming up.

For those of you participating via webcast, you can email your questions to alternativescores@FTC.gov, Tweet it to #FTCPriv, or post it to the FTC's Facebook page in the Workshop Status thread. Please understand that we may not be able to get to all your questions today, but we will definitely try.

MS. ARMSTRONG: I'm Katherine Armstrong and thank you for coming here today and thank you to our panel.

There's a lot of buzz these days about data brokers and alternative scores that are used to predict consumer behavior, such as the likelihood that a person will be interested in a specific product or service or that a particular transaction could result in fraud, but we want to try to get past the buzz today.

And so we have a panel of experts with different perspectives, different experiences, and our goal is to learn more about what is happening in the alternative scoring space, what may be on the horizon, and the privacy and consumer rights concerns that these products may raise. Our focus today is on non-FCRA covered products, although we do know there is a robust conversation to happen there.

Before we introduce our panel and start the conversation, however, Claudia Perlich will give us an overview of predictive analytics. Basically, a high-level, nuts-and-bolts presentation about how scores are created. After that, we would like to spend the first half of our time discussing the various kinds of products available, their uses and accuracy issues.

Then, Ashkan Soltani will give us a brief presentation about the trends in online pricing. And then, the second half, we will focus on the issues involving privacy, future uses, and the regulatory landscape.

But first let me introduce Claudia Perlich, who is the chief scientist at Dstillery and concentrates on data analytics for companies, real-world applications, and she also teaches data mining for business intelligence at the NYU Stern MBA program. Claudia?

OVERVIEW OF PREDICTIVE ANALYTICS

MS. PERLICH: Thank you very much. So thanks for having me here. It's a slightly different audience from who I usually talk to.

So I'm trying to get to the basics of the technology of predictive modeling and I have a couple of examples of that. So ultimately, what it means when we're talking about predictive modeling that's underlying a lot of the questions you're interested in, is algorithms that learn from data.

So you just give it the data and you let the computer do its thing and it comes back with, ultimately, models that affect your everyday life in some way.

I've picked two examples specifically to kind of explain initially how that works, and my first example is Lending Club. It's a microloan site where you can apply for loans and where investors can go and choose what loans they want to fund based on the information provided. And they actually provide the data, and I want to take this, initially, to just give you a very high-level overview of what might be happening under the hood here.

So let's just look at a simple example of loan default. Let's say I want to predict the probability of loan default. I can do that in order to make decisions whether or not to invest, I can do that in order to decide what interest rate is appropriate, I can even do that in order to increase the probability of my loan getting funded. There are many different reasons why I may want to know that.

So let's look at this. This is data where I just give you two variables here. One is the age of the person applying and the second one is the income. Now, I also have historical outcomes. That's key for all of this technology, that you actually need to have observed some of these things and what happened.

So let's say for those loans, we actually know who defaulted and who didn't. What happens next is, I give it to the algorithm. And I'm just showing it to you here in two dimensions to keep it simple. You look at this information and you have those people who defaulted and those who didn't.

What the algorithm is trying to do, without any human intervention, is can it explain the data. Can it find out what's different about the loans that defaulted versus those that didn't. And one technique, called decision trees, starts splitting the data along different dimensions.

And the first thing it might see is, well, if you look at the income, it says balance here, that's a clear separation that already gets a lot of the same on one side versus the other. And then it continues to do that by asking those kind of questions, how do I split it. Again, all of this is done automatically. There is an algorithm that does it, that tries all the different ways of splitting, evaluates which ones have more defaults in a clean bucket versus not, and goes ahead.

Now, at the end of this, this is called a classification tree, what can you do with it? Well, you can ask for a new applicant, what the tree thinks the most likely outcome is going to be. And the tree will just fit it into the bucket and it will say it fell there, and therefore the probability of default, given what it has seen so far, is roughly four out of seven.

So that's just one algorithm that uses data that you can then translate into a model. And you can ask it for a new case, where you don't know the outcome yet, what the model thinks is the most likely scenario.

A similar example here, it's a different algorithm that does the same thing. So the one thing to know, it's not like one algorithm we're talking about. There are hundreds of them out there and often they are just slightly different from each other. It does the same thing, except that it doesn't try to split the data, but it's trying to find a line that optimally separates them.

Ultimately, what it means mathematically, and that's the only equation I'm going to bother you with here, is it's trying to estimate these betas that you see there. So once you know the betas, and it's not moving forward here, if you estimate the betas, you can then ask the equation, what does this model think is the most likely probability of default.

So the takeaway here is, we use data where we have historical outcomes to build models that then we can ask, when we don't know the outcome yet, what is the most likely scenario. Now, one thing I wanted to make clear here, this is a version where you see it, the model doesn't understand what the axis means. The model doesn't know that what one is H, it doesn't know that the other one is income. The computer actually doesn't care whatsoever, it's completely agnostic about what is going on underneath. It just solves the one question I ask it, what is the best separation you can come up with and I have to specify best in some way.

Now, let's look at the Lending Club data that is publicly available. You can go there and download the information, if you want to. And they have a lot of information up on the website, including the textural description of the loan, categories, demographic information, in addition to credit scores. All of that is available, if I wanted to build this model. And here is the pull I did yesterday, so you can get that data. It doesn't have names in there, but it does have zip codes of where the person is from.

Now, I want to move on to what I do, that's my day job here, we're talking now about targeted online display advertising. So it's about all these pesky little ads showing up whenever you're surfing the internet, all over the place, trying to make you buy stuff. What is the data and how does this work?

So this is me and my company. And if you are browsing the internet, and you don't have third-party cookies disabled, and whoever wants to know the technical details, I'd be happy to talk about this later on, what happens is you will come to certain sites that have data partnerships with many, many, many constituents, including us. What happens, for instance, if you read a blog, if we have that data partnership, there is a forwarding of that and we can put a cookie on your computer.

Now, there is a lot of misunderstanding of what a cookie actually is and I'll talk a little bit about what that cookie does. It basically just contains a 20-digit random number that we have assigned to you. The moment you delete your cookies, it's gone. If you disable third-party cookies, it never even gets to your computer. If you have it enabled, then I can put this little piece of information that is just stored. It's not a program, it's just a piece of information that I store on your computer.

Now, what happens next? We are running campaigns on behalf of marketers, so marketers have an agreement with us. They come to us and say, we would like you to show ads, display ads, on the internet on our behalf for our product. The first thing we will do, we will put another one of these pixels on the homepage of that brand, so now we also see who actually buys that stuff. So we'll see people who go to that brand's homepage.

One important thing is, I don't see what Amazon knows about you, I don't see what you do on eBay, I don't see what you do on Facebook. I get a very partial view of what you do through these data partnerships and it's very, very far from complete.

Now, once I have that, let's move on to the part where the actual display advertising happens. You continue surfing the internet and you get to a page that actually has space for a display ad. At this point, there will be a realtime auction through an ad exchange. As the page loads, the different placeholders for advertising are sold in realtime, through an auction. We are getting bid requests. So the publisher, say New York Times, forwards the bid request to the ad exchange, the ad exchange sends it to us, and many, many, many like us. There is not just one ad exchange, there are probably 20 or 50 of them. We have 30 milliseconds to decide whether or not we want to bid on the opportunity to show the ad.

The important thing here is, what I do know at the moment is that same 20-digit random number that I assigned your cookie, if you deleted it, I don't know it anymore, but -- so I don't know who you are, but I know you are the same you you used to be when I saw you before.

I make my choice, I bid, and if I win, I get to show the ad on behalf of one of the 300-plus marketers who are working with us. After that, we are basically looking for post-view conversion. So we are not looking who is clicking on that stuff, but we actually want to see, is the person afterwards going to go to that brand's homepage to either buy it or at least check out the product.

So that's kind of, on a very high-level, and not talking about a lot of technical details, how that thing works. The predictive modeling piece that we talked about comes in many different places. The core question I have to answer is, who is even interested in that brand or in running shoes or in dog food?

So the first question is, can I build a model that predicts how likely it is that you are a runner in the first place, how likely it is that you are interested in dog food? So that's the first piece I need to solve, but there are many more. There is when should I advertise, how much should I bid. By the way, the creative, I have no control over. The brand just gives me that thing and says that's what you have to show, so that doesn't fall into my responsibility here.

A good question is, do these ads actually do anything. We can also decide what data we need, what data we buy, because we actually have to pay for that, what is the quality of data, and there is this notion of attribution that is more about kind of the cost incentives on the other side.

So there are many problems that ultimately rely on predictive modeling to be answered just alone and in this one application.

We also have issues around fraud, trying to decide which of these bid requests are real people as compared to bots posing as people and kind of malware systems that overtake people's computers, posing as them, when in reality the person may actually be doing something completely different.

What I want to highlight here is what is the data we actually see and collect, because that's important in the discussion you're having. So ultimately, what I see is a partial browsing history of you, only from those data partnerships and also from the bid requests. I'm not interested in understanding what you're reading on the internet, I couldn't care less. I'm going to take the URL that you go to and I hash it into a random string.

So you see, on the right side of the slide, there are two parts. There is the browsing history, it's kind of a timeline of hashed URLs. You have a random number that is kind of your ID that means nothing to anybody but me, because I use it to kind of append things to your history. And I also get, from the brands that we work with, these kind of purchase events. When you actually went to buy a product from one of the brands that we work with, because that's the thing I need to predict, right? That's ultimately what I'm interested in.

Now, the interesting point here, I do not want to understand you. I don't care to know who you are and what you do. I don't want any PI information for that. It's purely agnostic and translated into something that machines can work with that nobody else really gets to.

Now what happens next, when I showed you a model that I estimated in two dimensions, I am now going to do the exact same thing, but in 10 million dimensions. Every single URL is kind of its own dimension, did you go there or not. And we have roughly on the order of 10 million of these URLs that see over the period of time.

So I'm building this model for every single product that we are advertising to for the marketer and then I use it to score people. What means scoring people is, again, I see a cookie coming in, I look at the history I have observed. For that cookie ID, the 20-digit numbers, I estimate -- I ask the model what it tells me, it gives me a probability. Once you reach a certain highlevel, you become the target bucket. And whenever anybody from that top tier, typically it's like one percent, comes in a bid request, we decide to bid for that person. This is then downhill.

So I just have a short list here in terms of all the different places where this is happening. Spam has been doing this for a very, very long time. There's a lot of fraud and fraud detection, financial trading industry, a lot of work on this in medical diagnosis and quality control, sentiment analysis for blogs works exactly the same way. You look at the text and try to decide is this a happy or an unhappy person, or is that text from a happy or unhappy person, and many, many more, one of them being advertising and targeting.

I'll skip the next slide, because I think we'll just -- well, I guess the point, from my perspective here, it actually doesn't matter that much what the exact algorithm is. There are almost all kinds of equivalents, some work better sometimes, but at the end of the day, any of them will, more or less, do the same thing.

What matters much more is the data you feed it. The behavior of the model can only be explained if you understand the data that went in to determine that function. The algorithm itself is just a translator, if you want.

Quality control is incredibly hard. If you ask me how good my model is, I have no idea and I built it. I don't know whether working with it another week will improve performance by 1 percent or 10 percent, I don't know that either. So at the end, that's kind of my skill and intuition. You have this problem where models really are the skill of the person who assembled the data, and that's only as good as it is.

And finally, it is extremely difficult to understand the nuts-and-bolts of what the data is and how it affects the outcome. It's a really complicated problem, even for people who do that for a living. And with that, I'll leave it.

PANEL DISCUSSION

MS. ARIAS: Great. Thank you, Claudia. Wasn't that a fantastic presentation? Yeah.

All right. So before we begin, I would like to introduce everybody on our panel. We have a very, very long list of panelists, so I obviously will not be able to cover their fabulous information, okay? But I will try to briefly give you an overview of what they do. And make sure you look at the agendas that are behind -- you have the bios of everybody behind the agenda and so you can take a look at their fabulous, fabulous history.

All right. So first, next to Katherine, we have Rachel Thomas. Rachel is the Executive Director of the Data-Driven Marketing Institute and Vice President of Government Affairs for the Direct Marketing Association. She not only conducts independent academic research regarding how the responsible use of consumer data shapes industry and society, but she also represents the data-driven marketing community's policy-making interests on Capitol Hill. Thank you for being here with us today.

MS. THOMAS: Thanks for having me.

MS. ARIAS: Next, we have Stuart Pratt. Stuart is the President and CEO of the Consumer Data

Industry Association. He not only represents businesses that provide companies with the data and analytical tools necessary to manage risk, but he also has advised U.S. Presidential and Gubernatorial task forces on the importance of the freeflow of information to the U.S. economy and he testifies regularly before Congress. Thank you for being with us today.

Next to Stuart we have Ed Mierzwinski. Ed is the Consumer Program Director and Senior Fellow at the U.S. Public Interest Research Group. He often lectures and testifies before Congress on a wide range of consumer issues, including privacy, and he recently published a law review article on alternative scoring products through the Suffolk Law Review. Thank you, Ed, for being with us today.

Following Ed we have Pamela Dixon. Pam is the founder of the World Privacy Forum. She not only has written numerous studies on privacy, but she also has testified before Congress and Federal agencies on these issues. She has a study on alternative scoring coming out very soon, I've been told, so make sure you keep checking out the World Privacy Forum's website. Thank you for being with us, Pam.

Following Pam, we have Joseph Turow. Joe is the Robert Lewis Shayon Professor of Communication at the University of Pennsylvania's Annenberg School for Communication and he has published multiple books and articles relating to mass media industries. Thank you, Joe, for being with us today.

Next, we have Claudia Perlich, who we introduced before. And she gave us that fabulous presentation to start us off today. Thank you for being with us, Claudia.

And finally, to complete this very great panel, we have Ashkan Soltani, who will be giving us a presentation later on today. Ashkan is an independent researcher and consultant focused on privacy, security, and behavioral economics. He previously worked here at the Federal Trade Commission and he was the primary technical consultant on the Wall Street Journal's "What They Know" investigative series. Thanks for being with us today.

All right. So with that, let's go ahead and begin. And I would like to start today with Rachel and Stuart. Why don't you tell us a little bit about the history of these products. How exactly did they come about, alternative scoring products?

MS. THOMAS: You want me to start?

MS. ARIAS: Sure.

MS. THOMAS: Okay. Can you guys hear me? Are these on?

MS. ARIAS: Yes, if you -- I'd like to remind everybody to make sure you speak into the microphones. I know they're a little sensitive, but otherwise the folks in the webcast won't be able to hear us.

MS. THOMAS: Great. Thank you, Andi and Katherine, and good morning. Lovely to see all of your faces.

So I'm going to talk about marketing analytics, predictive analytics, similar to what Claudia introduced to us, because that's really a much better term to describe really what's going on in the marketing world.

So let's start with, the goal of marketing in every case is to meet consumers where they are with an offer for a product or a service or a cause that they might be interested in that is going to be of interest to them. So predictive analytics, no surprise, predict a consumer's likelihood or propensity to be interested in that particular product or service. That's the goal.

Now, of course, everybody here knows everybody gets marketing offers. The difference with predictive analytics is that those offers are more likely to be actually valuable to the consumer or the donor or the potential voter. So a consumer's propensity, of course, to buy a particular item is always changing. If I bought a car yesterday, I'm probably not going to buy one tomorrow. So those predictions about the future are constantly changing as well.

So what markers are interested in always is that extremely dynamic set of interests that a consumer has from day-to-day. Not just at any given point, but also in the different contexts, whether online or in a store, et cetera.

Now, it's important to recognize -- Claudia talked about the latest-and-greatest, but businesses and others have been using predictive analytics for more than a hundred years. Back in 1888, when Sears was getting started with their first catalog, they made the very smart prediction that folks living out in the rural west were going to be more interested in a catalog of consumer products than folks that had access to a lot of stores nearby and could walk in and buy them themselves. So they focused their marketing in the rural west, instead of those of us hanging out on the east coast.

Similarly, in 1912, L.L. Bean made another smart prediction that folks who had hunting licenses in

Maine, but lived outside of the state of Maine, were probably going to be interested in a catalog with hunting goods in it. And so again, that was how L.L. Bean got started, with a very smart prediction and the purchase of a list of folks with out-of-state hunting licenses from the state of Maine.

So fast-forwarding back to today, Microsoft had some really interesting research that came out just a few months ago talking about how, in asking consumers what they're looking for, they want more personalization, not less. Not just in the offers they get, but a seamless experience, whether they're in a store, or on an online site, or even in a mobile version of a retailer's site. They want to be understood throughout that whole purchase journey. They don't want gaps between all of those experiences.

So to meet those kinds of fast-moving and very personalized expectations that consumers have, marketers use those predictive analytics to make sure that they meet the customer wherever he or she is, with what they're most interested in, and however they're most interested in engaging. Whatever different context.

So for example, today in a department store, the store might look at what a customer has bought in the past, different products from different departments of that store, and look at its larger purchase history to say what other customers have bought those products and been interested in other things that a customer hasn't yet purchased. So they are going to analyze that and compare and, using those predictive analytics, they're going to guess. They are going to guess whether you were more likely to be interested in the coupon for the jewelry department or the kitchen appliances department, because maybe you bought that car and now it's time for a refrigerator, or apparel, et cetera, et cetera, et cetera. So that's business.

Nonprofits, you may or may not realize, use very, very important uses of predictive analytics to keep fundraising costs down by focusing on the people most likely to donate. But also to hone in on populations in greatest need of assistance and tailor their outreach to those populations, to make sure that they're most easily able to engage those in greatest need.

The Humane Society and World Vision recently have upped their ante in terms of targeted fundraising. They have actually created statistical profiles, not dissimilar to what Claudia was talking about, of their major donors so that they can then go out in the marketplace and look for others that fit those profiles of folks likely to make large donations to their organizations.

Political campaigns as well, incredibly important users of predictive analytics, to target political advertisements, whether in the mail, online, in realtime. Pandora has a great new service coming out that will let candidates or political organizations target those of us who spend our days with Pandora on in the background as we're working on whatever, this and that.

So how are they going to do that? They're going to look at public data of who won what elections, in terms of candidates in different zip codes. And then, you know, when you sign in to Pandora, you put in your zip code, they are going to see who listens to what music in those zip codes. So when a song that has been identified as, perhaps, making a correlation to an interest in a particular party comes on, you're going to get an ad for that candidate or that party as well. It's as simple as that.

So in all these cases, the organization is looking at, it's analyzing information that it already has about its customer or its donor or its voter to understand and make a best guess to predict what else that individual is likely to be interested in. Sometimes they can make these often, they can make these predictions just by analyzing the information that they have themselves, maybe having third-parties help them with the analytical power, like Claudia's company and others do.

Sometimes they might need additional information in order to make that leap to the next type of prediction. So they might go to a third party, a marketing information service provider, some sort of a company, who can help them with the analysis and with additional information to figure out what that customer might want next.

So predictive analytics are important, not just for keeping your existing customers and donors happy, but for finding prospective customers and donors and voters as well.

So if a company, for example, knows that customers are most likely to buy navy blue suits if they are women of a certain age in a large urban area. So if they know that, they might go to a marketing information service provider and say, the women who like the navy blue suits, what else are those folks fitting in that demographic likely to buy. And they're going to find out that a pair of nude heels, nude-colored heels, is going to the perfect thing to go with that suit. And they should serve a coupon for that, instead of a purple set of shoes, for example. So very important decisions affecting our daily lives. Mine, at least.

So some of you may be asking yourselves, what is new here? This isn't surprising, or maybe it is, but what's new? The predictive analytics obviously are not new. As Claudia rightly described, what's new is the analytic technology that helps get the predictions right. What's new is the power to actually get it right and give an offer that is of interest to you, as opposed to the person next to you.

So taking a step back again, whether this is your bread-and-butter, whether this seems shocking or magical or completely mundane, at the bottom line, it's really important to remember what this is all being used to accomplish, relevant marketing. And that's it. This is marketing data being used only for marketing purposes, and I'm happy to talk more later about how DMA makes sure that that's true, to predict the likelihood of a consumer being interested in a certain product or service over another.

Marketing data is not used to determine that individual's eligibility to receive a product, like a financial product or an insurance product, and it's certainly not used in any other kind of eligibility decision either.

So at the end of the day, for better or for worse, the biggest impact that marketing analytics will have on a consumer's life is whether or not that individual gets an ad or an offer that is relevant to her interests or one that is not. And we would argue that the proof that predictive analytics are valuable to consumers is in the consumer reaction when they do get a relevant ad.

Andi mentioned the research that I worked with folks at Harvard and Columbia on, on looking at the value of data. And in this area of marketing analytics and the flow of marketing data, we found that, in 2012 alone, $156 billion were added to the U.S. economy, 675,000 jobs in the U.S. alone, and 70 percent of that value was derived by that flow of data being used for analytics between first and third parties in responsible ways.

So we would argue that that's a value worth preserving and one that is incredibly important to getting it right for customers.

MS. ARIAS: Before we go and jump maybe into what Stu and Pam may want to say --

MS. THOMAS: Sure.

MS. ARIAS: -- about the uses, I wanted to give Stu an opportunity to maybe talk about other uses, besides just marketing, that these alternative scores or predictive analytics are being used for.

MR. PRATT: Okay. Well first of all, thank you all for inviting CDIA to be on the panel. It is good to be here. I like the word fabulous. I've decided I'm going to use that in some of my other presentations.

MS. ARIAS: You are fabulous.

MR. PRATT: And Ed and I always feel -- you know, we're on a lot of panels together. So Ed, I think in the future, that will be our theme, fabulous.

MR. MIERZWINSKI: And I'm fabulous. But note, I'm to his left.

MR. PRATT: Right. And appropriately so, I think. So even the seating charts are worked out just right.

So I'm precocious. My family would disagree, and I say it often to them, I'm precocious, and they continue to ignore that, particularly my sons. But I didn't realize that I was a big data analytics guy when I was a child, so I want you to know this.

I lived overseas and the only way that we knew what we wanted was when we waited for the catalogs from the United States to arrive in the 1960s at our house. And so our parents, one year, asked my brother and me, well, what do you want for Christmas? And we decided the best way to present our wants, our many wants, many, many wants, was to tear pages out of the catalogs from the various retailers.

And then Jim and I actually went into their bedroom and taped them to the ceiling of their bedroom. So that was just-in-time advertising, delivered at just the right time, with big data analytics, right down to the level of our needs, so that our parents could meet our needs in a way that they otherwise wouldn't have been able to do so.

So I just want you to know that big data has been around a long time, including in this very sophisticated way that my brother and I pioneered many years ago. If only I had known that, I probably wouldn't be working here today, I'd be retired on some coast and looking at oceans.

So we're -- CDIA, Consumer Data Industry Association, we really work with another part of U.S. dataflows. We work with companies that are aggregating data to manage risk. And risk matters, it matters a lot. At one time -- I used to have a harder time convincing folks of this, but if you just say two words, Great Recession, we all kind of get it. Risk matters, in a lot of different ways.

In the 1990s, risk was mostly focused on what we call prudential risk, you know, how do you make a lending decision, how do you make sure that banks are safe and sound. But most folks shrugged their shoulders and said banks are always going to be safe and sound. They all seem to be doing pretty well, bricks-and-mortar looks okay. But in the late ‘90s, we began to see identity theft cycle up and we realized there were different risks, risks that had to do with whether or not we actually knew the consumer with whom we were doing business. And because of the internet, which really was something that began for most of corporate America in, you know, the early 1990s, flowing into the next millennium, it began to also be many, many more transactions that were essentially where the consumer was remote to the transaction. They didn't know who that consumer was.

So my job here today is a little bit complex though, because the majority of the transactions we talk about, the majority of the dataflows that we represent, are regulated under a variety of different laws. And Katherine has assured me that I'll get pulled off the panel if I spend too much time talking about all those laws, but I'm going to do that just a little bit, just a little bit.

But our members’ dataflows do a couple of things really well. They encourage competition, for us as consumers. And that's good for us as consumers, more offers, different offers. It gives us a chance to evaluate different offers. Those offers can be at our desktop, those offers can be delivered in a variety of different ways.

It also -- our data is really a framework of safety wrapped around the U.S. economy. Fraud prevention is elemental in a lot of different ways. It's identifying the consumer in a card-not-present transaction. It could be as simple, by the way, as a retailer who doesn't have a bricks-and-mortar operation trying to understand whether or not an address to which they're sending a very expensive item is or is not zoned residentially.

It could be devices today, and device recognition strategies, to try to understand whether or not I, with the -- if I'm a business and I have a current, ongoing relationship with a consumer, whether or not I recognize the device that the consumer is using as he or she engages in these transactions.

So there's layers and layers of fraud prevention that occur. It's all seamless. We don't see it, we don't feel it, we don't think about it. We're only upset when we discover that we've become a victim of some type of crime.

Nine billion times year a our members’ data is used in what we'll call a risk transaction of some sort in the United States. Our members are also the largest global companies, delivering and propagating these same types of services around the world, to some of the fastest developing economies, economies like Brazil and India and so on. They know what they're doing.

They are managers of big data. They are managers of big databases of data. It's primarily structured information, though. Sometimes it's hard to know what the definition of big data really is, but the kind of data that our members are gathering could be derived from fairly sort of pedestrian sources. It could be public record data gathered in the United States that could be used, which helps us with mortgage frauds and flipping and issues that have to do with the safety and soundness of the mortgage that's applied to a property.

And like I said before, it could be a database of known fraudulent applications that have been pooled by various retailers or other transactors in the marketplace.

If there's a dividing line between Rachel and me, though, and it's a great symbiotic dividing line, there's a baton pass. Our members benefit tremendously from the fact that there is this robust, incredible, targeted system that connects consumers with what they want. And in America, that's okay. We like buying things. We like engaging in the marketplace. We like seeing that offer that makes sense to us.

Then, there's a baton pass. A consumer chooses to click and a consumer chooses to apply for something. And that's more often where our members then kick-in to the process. More often where -- and in fact, if you're in the financial services space, more often where you are complying with laws like Section 326 of the U.S. Patriot Act, know your customer. Red flags rules, promulgated by government agencies such as the Federal Trade Commission, are properly protecting consumers against identity theft.

So there's this, again, this confluence of dataflows that occur on the front end of an application, which is occurring sometime after I've seen banner ads or I've shopped in bricks-and-mortar stores or done whatever I do as a consumer to figure out what it is I want and at what price I want it and so on and so forth.

But some of those laws that regulate our industry, Fair Credit Reporting Act, Driver's Privacy Protection Act, the Gramm-Leach-Bliley Act, Title 5, again, not the topic of today, but important for you to know that these laws wrap around a lot of these different databases that are out there for consumers. Because these databases particularly -- and Rachel used an important term, eligibility. Once you cross over the line into eligibility, one you cross over into what I call gate-keeping, you get a yes or no. Or the yes that you get is the best yes on the list or it's a qualified yes, somewhere down on the list, you pay us a little higher price or a lower price, all of that is regulated under a variety of, for example, fair lending laws, the Equal Credit Opportunity Act, the Truth in Lending Act, and so on and so forth. So once you get into the application context, you're back in that world of laws that wrap around the transaction.

So let me just give you a couple of examples. Lending Club, I want to go back to Lending Club. I don't know a lot about Lending Club, but I'm just saying that if Lending Club has a new, innovative way of managing big data to try to make a lending decision, even though Lending Club is using new data sets, they are still obligated to comply with the Equal Credit Opportunity Act. They are still obligated to comply with the Truth in Lending Act. They are still obligated to make sure that they don't have disparate impact problems. The confluence of all of these laws still applies. So there's nothing new about the use of that type of information, when it's in the context of that type of application process.

And then the different example. The CFPB was looking an annualcreditreport.com, by the way. By the way, let me say that again, get your free report every year at annualcreditreport.com. It's my little advertisement, but not a bad one, right? Not a bad one.

So 16 million consumers roughly are looking at free credit reports each year, out of 200 hundred million-plus consumers, so it's not a huge amount. So the kinds of analytics that have been discussed by Claudia, and also by Rachel, might be one way for us to reach out into that community of consumers more effectively and try to find those consumers who we think would benefit from accessing free credit reports, but aren't doing it today.

So there's, I guess, a social good example of how there's a nexus between get your free report, be credit report literate, make sure that you understand what's in your credit report, you know, all of those things that we believe in, and the analytical tools that can allow us to kind of get to that point and reach those consumers who we think are most likely to, you know, point-and-click and move forward.

So that's just an example of the power of this kind of information and how it connects consumers with, sometimes, maybe something they don't know that they should do, but they get it.

MS. ARMSTRONG: And that's a great example, Stuart.

As Andi and I are mindful of the clock, we'd like to turn quickly to Ed and Pam to see if they could mention some other products, what they might be used for, and then before Ashkan's presentation, we want to talk about data accuracy.

So I want to make sure we get to that before the top of the hour.

MR. MIERZWINSKI: Well thank you, Katherine and Andi. My work at U.S. PIRG is as a consumer advocate. And I'm concerned not about the data, per se, I don't think anybody is. I'm concerned about its use and its impact on financial opportunity.

And I'm also going to give a disclaimer, as Stuart did, that I'm not here to talk about the Fair Credit Reporting Act, but I have to mention it peripherally, or at least in passing.

For 40 years, financial marketing of the most import kind, based on your credit report, the most detailed profile about you, has been governed by the prescreening rules of that Fair Credit Reporting Act. That law says, if a company wants to use your detailed financial profile to market to you, it can only market to you for credit or insurance purposes, not direct marketing, it must give you a firm offer of credit, and you have the right to say no to that use of your information. You have the right to say no to using your financial profile for marketing to you. And the kind of marketing that can be done is extremely limited.

I am very concerned that we are moving to a new system of unregulated, wild west companies, running roughshod over consumer rights on the internet and making decisions about what ads to serve to you, maybe not directly determining eligibility yet, but deciding what box to put you in, what place to direct you from your cookies and from the other information that they have about you, and possibly causing you to pay more or get fewer opportunities than other consumers. That's the short version, I've got much more to say.

And by the way, it's USPIRG, P-I-R-G. If you go to my blog, on the homepage today there's a lot more detail and links to some of our materials, including my paper with Jeff Chester at the Suffolk University Law Review.

MS. DIXON: Good morning. Thank you so much for the invitation, I really appreciate it. This is a great panel and I really appreciate the opportunity to share a discussion about this important topic.

So let's begin with the fact patterns here. So the first fact pattern is that scores are proliferating. In the past, when the credit score was developed, the credit score used limited factors, well under 100, they were controlled factors. In fact, those factors that are used in the credit report are regulated. They cannot be discriminatory, they cannot be prejudicial. And Congress did this for a very good reason, the same kinds of reasons that they passed the Civil Rights Act. There should not be any kind of hidden discriminatory factors in scores, this we can all accept as a baseline. So that's one thing.

The second thing is that the large data set world that we're living in is not going to reverse itself somehow. That genie is well out of the bottle. So given that, really one of the ways that all of us make sense of our world is by shortcutting. Understanding data and predictive analytics allows us to do that, the machines do the hard work of sifting through petabytes of data for us. So the results of that are spit out are often scores. Scores can have varying ranges, they can have varying values, and can mean completely different things, depending on the factors that are fed into the score, as Claudia discussed, the algorithm, and then, of course, we're talking about the use of the score.

The really important thing here is that the credit score had a very focused purpose. Today, with the real proliferation of the technologies that allow more and more retail and enterprise and small businesses to create predictive analytic scores and tools and results, it's becoming more important to find out what other scores are out there. And that's the second fact pattern, there's a lot. So that's the second thing.

The third thing is this, the credit score, as a controlled score, has been very, very carefully observed and has a lot of oversight. The new scores don't enjoy that same kind of protection. So here's my thinking on this. We really need to understand that there is a continuum of scores here. Not all scores are bad. In fact, some scores are actually helpful. The Equal Credit Opportunity Act mentions specifically credit scores and how they can assist in reducing discrimination in lending. This is a good use of a score and it's a regulated use of a score and it's appropriate.

So today, we really need something like that to look at these scores that have proliferated and are new, so let's talk about some specifics. The credit score, a few factors, and a static score, doesn't change that often and that much, unless you really game the system. And that's a whole different matter. Not for this panel, right?

If you take an aggregate credit score, however, an aggregate credit score and actually some modeled credit scores, can use 1500 factors. These factors are in a big black box, we don't know what those factors are, we're not told what the factors are, and yet Claudia's presentation was completely correct when she said, look, you have the factors that go into a score, really that's everything. Good factors in, nondiscriminatory factors in, much better chances of getting a nondiscriminatory score on the backend.

But if there are credit-related or any kind of eligibility-related scores that have discriminatory or prohibitive factors that are used in the score soup, we've got a big problem. But we won't know that we have a problem, because right now most scores are secret, with the exception of what I would call social scores like Klout. Consumers don't have the opportunity of learning about the scores, because there is no transparency for them and certainly the factors are secret. So we've got big problem there.

Now having said that, there's something really important to understand. And I'd like to echo Ed's remarks. We did a thought experiment and we asked ourselves, could the Klout score, a social-influenced score, be covered by the Fair Credit Reporting Act. And you just can't get there. There's a real first amendment issue here that we have to grapple with. There is such a thing as free speech. And you know, if someone is quoted in the Washington Post, and the quote happens to be not so great, right, and it makes the person look bad and they don't get a job because of that quote, does that mean the Washington Post should be regulated? No. None of us think that, right?

So we have to be really careful here. It's attention.

MS. ARMSTRONG: Pam, that's an excellent point. And I want to save some of this conversation about the parameters until the second half.

And although I have a very specific thing I want Joe to share with us, I think that right now, this is an excellent segue into the accuracy issue. And as Claudia mentioned, quality is hard. And as we look at it, accuracy has two components. One is going to be the model and the other is going to be the data.

So I'm wondering if anybody could speak to how companies determine whether the data -- well, whether there are certain sets of data that are inherently more accurate than others and is data accuracy relevant for all types of scores? So I'd like to throw that out for a few minutes, if anyone wants to comment on that.

MS. DIXON: Can I just jump in?

MS. ARMSTRONG: Sure, please. Absolutely.

MS. DIXON: Data -- scores are coming from public data, they are coming from demographic data, enterprise, social, even some health data. There is even financial interest and activity. So the accuracy of data is a huge issue. It's very, very difficult to create a score above 95 percent, everyone knows that. I think that, from the analysts I've talked to, a score above 85 percent is just awesome.

But I think that there's really no way for anyone who uses thousands of factors in a score to completely assure that each factor is accurate. I just don't see it.

Now hopefully, there will be a lot more transparency in the industry and we can find out a lot more about this. And that's what's incredibly important.

MS. THOMAS: Can I add to that?

MS. ARMSTRONG: Sure, absolutely.

MS. THOMAS: So I think when we think about accuracy first, again, from the marketing perspective, you want to have an accurate ability to predict what someone is going to be interested in. You are more likely to have a sale at the end of that, so yes, the data being accurate in order to make that prediction out of predicted analytics is a good thing.

That said, when it comes to consumer protection, I think it's incredibly important to look at the relationship between the importance of accuracy and the use of that data. Data is data is data.

So when we're talking about marketing, if data is incorrect, you're going to get an offer that isn't relevant to you and that's the end of it. If your credit score is inaccurate, you could be denied housing, insurance, et cetera, et cetera. Very important permissible uses under FCRA.

So I think it's important to play this out to the end. When we're talking about data, what is it being used for and, at the end of the day, that determines the importance of its being accurate, linked to the potential impact on a consumer, in terms of harm to their way of life.

MS. ARMSTRONG: Ed. Oh, I'm sorry. Ed and then Joe.

MR. MIERZWINSKI: I'll just be very brief. But I'll say this, how do you determine accuracy? Well, we need, as Pam said, more transparency. And there have been a number of studies by consumer groups where they have requested information about their profiles from various data brokers and others. And the profiles, when they've been provided, have been incomplete and inaccurate.

By the way, I'll just make one quick point. I've been a member of REI for 40 years, I've shopped there for 40 years, I've been a member -- not a member, but I've shopped at L.L. Bean for 40 years. And for some reason, L.L. Bean, for the first time ever, just sent this non-fisherman the fisherman's catalog. Maybe they're trying to expand my horizons? I don't know. But I thought that was pretty interesting.

MR. TUROW: I also wanted to add about the question of accuracy. In huge models, from what Claudia was saying and other things that I seem to know, it is very difficult to know what about the model is accurate or not accurate when you're predicting something.

And the other thing I wanted to say, connected to this, is HIPAA and Gramm-Leach-Bliley aside, there are lots of places to get data like that, that you can go around these laws.

There is, for example, a company called Medix which has a website that will give you discount coupons on serious health problem medications. So you go to that website and you write in what your problems are and then you get these discounts. And you don't know what they're going to do with those data. Privacy policies are fascinating obfuscatory. And it is very difficult to know.

I think we have to expand the notion of -- the notion of a credit score and a data broker. Is Kroger a data broker? Now, reading Kroger's privacy policy, they seem to imply that they don't sell their data, okay? But is it not selling data if Kroger allows advertisers to put ads on sites which track people and then, through the cookies, essentially they are buying data that way from Krogers? It's a side-door data broker activity, I would argue.

And so we have to think a little more broadly, I would argue, about data brokers and about the ways in which companies try to get around some of the obvious laws about protection.

MS. ARMSTRONG: Just one second. I wanted -- before, Claudia, you talk about the accuracy thing, before we leave the products and their uses. Joe, if you could describe the KLM example that you shared with Andi and I when we spoke with you.

MR. TUROW: Yeah, I remember reading about this. It's not so much a big data issue, it's just simply the notion that people can find out, through the website, if you're going alone on a KLM flight, what kinds of interests other people have and then you can decide whether or not you want to sit next to that person. It's not clear to me that other people have -- presumably everybody has the right to say they want their interests put out there. If they didn't, that would be a fascinating question of predictive analytics.

But let me, as long as we're getting at this, say one thing that did happen to me that I think people would not have predicted. I was on a United Airlines flight coming from Wisconsin into O'Hare and the flight to Philadelphia was canceled. They told me to go to a customer service place and scan the boarding pass I had for the canceled flight. When I did that, it gave me a number. And to the right, there was a monitor that said, the amount of time that you will be taken to be served will relate to your status, your loyalty status, with the airlines.

Now I don't think that most people -- and I was very fortunate, I had a lot of miles. But there were these poor people sitting in the back, they weren't treated for a long time. And the implication of it is, some people will get flights and other people won't. So --

MS. ARMSTRONG: So we now have a marketing risk/credit and airlines?

MR. TUROW: Yeah. I guess I'm saying is, we don't know, when we give these data, what the implications are.

MS. DIXON: Can I jump in for just a second?

MS. ARMSTRONG: Please. Everybody can jump in quickly, and then we're going to move to Ashkan.

MS. DIXON: I'm most -- I'm not as concerned about ads as I am about eligibility issues. Our focus and our research in this area has been on eligibility uses of marketing and noncredit data outside of the FCRA and outside of HIPAA.

I understand that the ad information is very fascinating. It's fascinating to me, too, on a research level, but we're very concerned about the more impactful scores. And those are definitely use of health data as factors, use of scores in health, which is being done today, by the way, that include non-HIPAA information held outside of the medical establishment and also use of discriminatory factors in eligibility decisions, which is happening today.

MS. PERLICH: I wanted to quickly jump on the accuracy. I'll keep it -- the first one, there is an inherent tension between the ability to regulate and keep the model understandable, or the score, and the accuracy it can reach. My being able to add 100 or 1000 more factors into the model, I can double the accuracy and you will not understand it anymore. There is that tension, we have to be aware of it. It's simply the ability to predict is a function of the kind of variety of stuff you give it to start with.

My second comment is on data brokers and scores. One of the reasons we have the data, we use the data we use, which is what I call primary, I observe people's actions, is anything derived from it that I can buy from data brokers is absolutely awful, in terms of accuracy, from what I can tell. So all the scores, whether it's done demographic, that I can buy your gender, and typically people are both -- so my experience has been that the data that somebody else has derived somehow is really problematic. And at that point, it becomes completely useless.

MR. PRATT: So just a couple of quick -- I'll move through these like a list.

MS. ARMSTRONG: That's fine.

MR. PRATT: So just a quick thought about data brokers overall, and this is just because, at CDIA, we deal with that term quite a bit these days. It's a newish term and it's a really undefined term. So I just want to make clear that really, when you're thinking about data and analytics, and actually, Joe, I think you said it well, it could be a closed system of data that has been aggregated by a single entity, like Kroger, or a search engine, for example. And they may not be selling the data, but they're certainly inviting people in to make use of that data, to deliver the advertisement. So we don't see, you know, this sort of third party versus first party thing as particularly relevant to the data broker issue overall, but I just think it's important to lay that out and say, you know, that there's both.

With regard to credit versus noncredit, and this goes to Joe's point about getting around laws, nobody is getting around laws. And we may have a debate about transactions of the definitions of those transactions, and I think that's what Ed is kind of driving at, but nobody's getting around laws. If you're engaged in making a lending decision, you're regulated by the laws which regulate lenders. And if you're doing it outside of those laws, you are violating laws and Maneesha's team and others will go out and find you and investigate you and prosecute you for not complying with the financial services laws that apply today.

With regard to developing credit scores, I don't like the term credit scores versus this better term, analytics. Because credit scores makes everybody think about the credit score that may be based on a credit report, but that's not really the point. And by the way, I think a credit score developer would say, it doesn't matter how many factors I have, I must build a score that is successful in the marketplace. It must be saleable to somebody who sees the outcome, sees what happens.

So if I'm developing it for marketing, I want to see that people click on the ads and engage and make purchasing decisions, because it's a good analytical tool that leads me to some place where I want to be to make a yes decision, because I like the yes decision I'm making as a consumer.

Credit is a little bit different. It's about prudential lending, of course. And so regardless of whether it’s 100 or 1,000 different factors, it has to be statistically sound, it has to be empirically derived, and whether or not it's a first-party score that's developed by the lender in house, based on big data, sort of the old Zest Cash model, or whether it's a score that has been developed by any one of CDIA’s members, who are some of the biggest data analytics companies in the country, whether it's a fraud score or whether it's a score for credit, the outcome is the key. It isn't measuring factor-by-factor precision.

But yeah, developers will look at a whole variety of factors. They may look at 100 times the number of factors that ever end up in the score, because they're trying to find the right sauce of scores that will lead to an excellent lending decision, allowing them to penetrate and get to more yeses than nos, but to do it in a prudential, safe and sound, sort of banking safety and soundness way.

MS. ARIAS: So Stuart --

MR. PRATT: So that was just some quick thoughts in all of that.

MS. ARIAS: I think you raise an excellent point. And it's actually a follow-up to you, Pam, because I think you brought this up and he's responding, right? What is an eligibility use that falls outside of the FCRA? What is it that you are envisioning in your comments?

MS. DIXON: Okay. So look, there are scores today that are called either aggregate credit scores or modeled credit scores, so let's talk about aggregate for a second. The Fair Credit Reporting Act, our lovely, trusty Fair Credit Reporting Act, applies to individual consumers. Aggregate credit scores apply to a neighborhood.

So if a person lives in a neighborhood, it's kind of like in old England where if you lived in a house where there was a debt collection, then you were also a bad apple, it's kind of the same idea. And I'm sorry, but how -- if you are using an aggregate credit score that is a very close proxy for a credit score, and offering a financial product or an insurance instrument to a consumer, I think technically they're not covered, right?

However, it's a really important eligibility decision or a product offering and this is where I think we have enormous tension. And these, I think, are the important scores to focus on and they are secret scores. I can't purchase my aggregate credit score, I can't. It's not regulated.

MS. ARIAS: So I know people want to respond. And I think the conversation is starting to move to the idea of what the effects of these scores are, right? So before we jump to effects, we really wanted to get Ashkan to give us his presentation, if he wants to come up.

So Ashkan is going to give us a brief presentation about some of the privacy issues associated with scoring products, especially about some of the research that he and other researchers have done on emerging trends in online pricing with the use of these scores.

MR. SOLTANI: Can I just briefly --

MS. ARIAS: Yep.

MR. SOLTANI: -- just make a comment on accuracy?

MS. ARIAS: Absolutely.

MR. SOLTANI: One thing I want -- and I'll talk about this a bit in my presentation, but I want to clarify that, with regards to accuracy and these traditional models, this stuff isn't as black and white as we are used to. It's not whether you're qualifying for -- well, some of it is, whether you are qualifying for a job or whether you are qualifying for credit.

But we should also be mindful of the ways these aggregate scores or these numbers are used to, what I want to call do fuzzy nudges, right? These are things like how long you wait line, is one of those things, where the score might have an effect on kind of subtle behaviors that kind of get up close to some of the things we care about. Other things are like how long you might wait at a call center. So call centers are actually profiling for customer service.

And I'll talk about this but, for example, what credit cards you are shown, what credit card objects you are shown and what you have an opportunity to apply to. You know, how far down on the page or whether it's on the next page, I think those things also influence. They're not direct kind of credit scoring type products, but they kind of influence -- we know most people don't go to the second page, right? We know most people will kind of see the options represented. And sure, you can definitely dig, but it essentially nudges people to a particular outcome that I think we start, you know, caring about.

And they might use explicit factors. They might not be -- they might be just data driven, right? So, Google, for example, is notoriously well-known for their HR recruitment algorithm. So when you submit a resume to Google, they have this awesome matching system that will kind of present you for potential jobs. And it's not clear whether -- and it's likely that they're using, say, sexual -- you know, sex or race as an indicator, but there might be that latent property that's in the system, that might emerge, that we care about. But it's essentially one of these fuzzy messages which is, ten resumes are presented to the recruiter and they are based on a bunch of factors that are essentially scoring, but not in a direct yes or no way, but just the probability of. And those things are a bit fuzzier, so I'll talk about a bit of those.

MS. ARIAS: Fabulous. Did you -- yeah.

MS. ARMSTRONG: So as Ashkan is setting up, I think I wanted to just observe that Pam and Ed have noted that we do have a statutory framework in the FCRA that covers some kinds of data.

And I think they're correct to point out that we're talking about these non-FCRA products, but using some of the lingo that the FCRA uses, and that's a bit of a challenge.

But now, Ashkan?

EMERGING TRENDS IN ONLINE PRICING

MR. SOLTANI: Sweet. All right. Oh, that's loud.

Everyone, so I want to just briefly talk about some of the research. So I want to just keep it, since it's a five-minute quick presentation, I'm just going to talk about some of the research that I did in the past on kind of scoring and use of data.

And so I'm going to briefly talk about the methodology and how I did it, the findings, and some of the -- some comments on data sources. And so most of this research was from a piece we did for the Wall Street Journal on sites varying prices and offers to consumers based on data about them.

And the methodology was really kind of basic. We kind of crawled the web with a variety of user-agents. This is, you know, your mobile device or your browser, it's known as user-agent. Sometimes Firefox is a user-agent, sometimes Safari means you're using an Apple product, usually. You know, whether you are using a mobile device, that's your user-agent.

We looked at different proxies for different locations in the world, like multiple locations in the U.S. and multiple locations in the world, and then we built profiles. We built essentially user profiles that would, say, browse, you know, (inaudible) sites and Scrabble kind of games versus, you know, potentially someone that only connects to the West Coast that checks car sites and electronics, right, to try to generate profiles. And we verified these profiles by looking at various dashboards for the profile managers.

And so we would essentially go through and iterate the web and check a particular website, say, 1,000 times or, you know, tens of thousands of times with the right permutations to see, statistically, what offers get offered, you know, presented to who. And so essentially kind of black-boxing or attempting to black-box some of these.

And so some of the basic findings. So user-agent, this is, you know, what browsers people use. Some people might have kind of read this article where, this wasn't our research, but it was another team, but they had identified how Orbitz was simply showing luxury hotels to Mac users, right?

So this was, how far down on the page or whether you have to go to the next page. And this is a subtle nudge. They argued that, well, look, like Mac users are posh, they want to spend a lot of money on hotels, so let's give it to them. There's no problem with that, this is just kind of a highlight. We actually found that, in fact, Orbitz was giving discounts, they'd call these Mobile Steals, to mobile device users, smartphone users, right?

So again, probably no problem with that. This was actually not Orbitz doing it, this was the hotels providing these offers through Orbitz. Orbitz provided them the functionality, the technology to do it.

But one thing to think about is, for example, who buys smartphones? Usually people who buy smartphones have some disposable income to pay for the data plan and pay for the $200 for the phone versus the free device. This was a few years ago. And so again, a good proxy for a loss leader for higher net worth individuals possibly, but also simply just using the user-agent as a proxy for this kind of thing.

I don't know if people fly, but in-flight -- probably one of my first experiences was when you fly, like on Virgin America, and you have in-flight wi-fi, the price they charge you is based on your user-agent. So this is both my desktop, but one advertises as an iPhone and the other advertises as a regular, you know, Firefox browser and they charge different prices.

And they might argue that you -- you know, maybe mobile phone users use less data than iPhone. Although both have an iPad video streaming app, for example, so it's not clear. But again, this is another example of like charging different prices. I don't know the legality of this, don't try this at home kind of thing, but it's a proxy for your device.

So location is another -- and Pam kind of raised a really good point about kind of location being an indicator for other factors, not just where in the world you are, right? So we looked at Staples, right? So Staples is the office supply store and they were charging different prices for goods based simply on your zip code.

And kind of we dug into a bit and it turns out that, in fact, the algorithm seem to have a higher likelihood to charge you more based on whether or not you were near, further or closer to a competitor's store. So they mapped out where the competitors were and essentially were charging you, you know -- in staplers, it was a dollar or two, but on some of the other items, it was up to $100 for the same safe. So you could order something online and, depending on where they thought you were, they would charge you a different price for the same good, right?

We also found that it wasn't just Staples.

Other suppliers, Home Depot, for example, Rosetta Stone, were giving you discounts based on where they thought you were, right? Where you were located. And again, even Discover, the kind of -- the credit card, was providing their it Card offers only to certain regions.

So if you tested the website, if you visited the website or if you visited sites that featured their product, they would only show that it Card to certain regions. The card had different benefits, different kind of deals, essentially, associated with it.

And what's interesting is we also looked at, for example, try to correlate the Staples stuff with weighted average income and, in fact, there was a correlation. So the places where people were getting charged more were, in fact, places -- places that people were being charged less were, in fact, places where people made more.

And then we're -- and so again, this is an interesting kind of commentary where it's good enough to provide offers or show you deals, you can always go to a store to see what the price is. But online, it was a little bit different because, in fact, all you have to do is, in this case, you didn't need a proxy, you could just set your cookie to whatever zip code you want to pretend to be, and they were relying on this inaccurate signal as a way to price you.

And then the last kind of important component of that methodology was kind of permutations and profiles. And I'll admit this was kind of one of the least kind of successful parts of our research, partially because it is very difficult to black box the profiles, as they change each time you sample them, you corrupt your profile each time you use it, right? So if I check a kids site as an adult, I might then attribute to those factors.

The other thing is the spam algorithms and the fraud detection algorithms are incredibly good, so they quickly identify whether you are a human or not and so they kind of -- the ad engines will kind of look for click fraud and so we were tripping up a bit against that. I have a new methodology of how I would do this next time, which is a bit more robust, kind of crowd-sourcing it essentially, or maybe more of like a botnet. But a consensual botnet. A consensual botnet.

But some of the things we found, for example, Nextag is a search engine where you can search for products, right? You can search for, again, I don't know why we were focusing on scanners and office supplies, but if you went through Nextag, you would get cookied as a Nextag kind of customer. And as such, you would get different offers, you know, up to $50 to $100 difference for the same items based on whether you had these cookies.

So again, profile-based kind of -- and they were doing really clever stuff to not trip up, not to let their competitors know that they were charging lower prices. So they would, in fact, show the price as a GIF, as an image, so that if you crawled their site, the price would be $499, a computer would think the price was $499, but the image that a human would see would be $350, right? And so they were trying to actually adjust and score differently on this site and then let users see the price. It was kind of some clever stuff they were doing.

And then finally, this is earlier work that one of my colleagues did and then we followed-up and verified the same results, which is CapitalOne credit cards. When you go to their site and try to pull up -- I think Emily actually worked on this story as well. When you kind of pull up a website and different credit card offers, you would get -- essentially CapitalOne would ping a data supplier, X+1, which is one of these kind of data networks, and they would score you as the type of customer you are, the type of credit you are likely to have. And those would, in effect, show what credit card offers you received.

And ultimately, of course, when you went through and applied for the card, the credit decision would hit a real credit check, right? Your qualifications of whether you qualify for one would be coming from, you know, Experian or one of these databases. But the offers you were given were based on this kind of fuzzy data that we're talking about.

And we found, you know, we found the same type of thing pulling data from Double-Click and RU4, which is X+1. This is still -- this was last night, or a couple of nights ago, I just verified that they still use these indicators. And to the degree that, you know, Claudia said, we all know what data -- some of these are actually explicit. So some of these they will explicitly ask for the types of sites or credit scorings and there is an API that specifies and a programming interface that specifies the type of customer that is.

So I've done some work with like some online dating apps and they specify what your marital status is or whether you smoke or whether you do drugs. And these are not fuzzy factors, these are explicit categorizations of what the data broker thinks. And they could be inaccurate, but they are at least specified explicitly, which is what's interesting. So whether you are high credit or medium credit.

And mostly, I want to just talk about the data sources, because this stuff is made possible by a number of front-end engines. This is Omnitures Test and Target Suite, all it is is it lets you kind of experiment with different offers, different profiles, and see what's the most optimum for whatever task you're trying to achieve.

And so again, nothing inherently wrong about the system. It's a very easy-to-use, quick-to-use system. You can vary the language and you can vary some of the attributes of, you know, the price, et cetera, but the key is that you can integrate this with a variety of data sources, right? So you can pull data, you can use for A/B this testing, a variety of factors.

This is, for example, geo-targeting or sex or age. And the question -- the thing that it raises is that you have these sophisticated tools that simply are just doing optimization, right? But they are pulling from a set of data, a large data set. Everyone has seen this graphic, it's like the most overused graphic, but it does indicate that almost everyone in this kind of red box is someone to purport to sell some sort of data or provide some sort of scoring that you can use to use this kind of A/B testing.

And so this stuff is going to have interesting outcomes that we've not anticipated, relying on inaccurate data without a lot of transparency, and very hard to black-box, as I've said. And so I think that's kind of the topic we want to dive into.

And just as a final note, we ordered the same -- we ordered the same stapler for two different prices and they both were the same. They worked the same.

PANEL DISCUSSION

MS. ARIAS: Thank you so much. All right, with that presentation, I really wanted to give the chance for the panelists, particularly Rachel and Stuart, I would like to give you a chance to kind of give us your thoughts about some of the kind of findings that Ashkan has found, in terms of the use of these predictive models and the effects that it might be having on consumers.

MR. PRATT: Do you want me to dive in or --

MS. THOMAS: Do you want to go first this time?

MR. PRATT: Okay. So I'm just going to quickly work my way through bullets and then Rachel can do an excellent job representing a much more detailed and probably better presentation of all of the details.

So a couple of things. First of all, I used to run retail stores all across the Washington area in another life, many, many years ago, and of course we rewarded loyalty. I mean, I don't know what the epiphany here is with -- you know, when I was on an airplane and they were pulling bags off of the airplane, you're darn right I wanted my gold card stuck to my bag, because I was loyal to that airline, to be the last bag they pull off the airline, because I'm hoping my loyalty means something.

So I don't know that that's an epiphany, that's been around forever. Everybody wants to sell, everybody wants loyalty, everybody wants to reward loyalty because you want more people to come back. So that's just kind of a macro thing. There's no ah-ha moment there. Everybody wants to be treated as a loyal buyer, that's all. I mean, again, whatever the point is.

With regard to differential pricing or whatever the term is, just a quick reaction to that as well. I mean, I go to the airline website and then I go to Kayak or some other aggregator and I look across both of them to see. So I guess the big message there is, you know, frustrate the analytics companies and go shop really aggressively and make sure that you understand the different pricing opportunities that are out there. And don't be linear in terms of how you behave online. So again, I mean, I think there is a -- it's not that there's an unsafe marketplace out there, but consumers certainly should shop.

With regard to the alternative data, other data that's out there, just remember that these new data sets are potentially going to allow us to reach consumers who are often not included, consumers who are sort of the unbanked and the under-banked, and to bring them into traditional lending contexts and others. And it may start with marketing offers before you ever get to firm offers under laws like the Fair Credit Reporting Act.

So big data is an opportunity for inclusion. It's got to be done right, the accuracy standards will apply when you move into the lending context. The fairness protections of ECOA and other -- Equal Credit Opportunity and other laws will apply. The protected classes of consumers are protected. But that's a really important point, that these data are opportunities to include, not just to exclude. I mean, I think so often this is a glass half-full kind of discussion. I think it's a full glass discussion, there's an enormous amount of opportunity here.

It's good to have a panel like this, I think, to flesh out a variety of views. I think it's a really helpful panel, I'm learning. It's good for me to be here, so I appreciate Andi and Katherine having me here. But I just think there's a better story. It isn't just a press down on consumers, but I'm still just sitting here as a manager of businesses which have relationships with consumers saying, you bet, I had sales people who definitely went to the people who bought the most first. Because that's how I got my profit per square foot and that's how I continue to pay salaries and monthly salaries that I had to meet. That's all there is to it.

MR. SOLTANI: Just a quick comment. I'm curious, who wants to be included in that higher priced consumer category?

MR. PRATT: That's why you shop.

MR. SOLTANI: Right, but I'm --

MR. PRATT: I mean, that's why you shop. That's why my wife is a much better shopper than me, she always has been. She berates me all the time because I would go to a big-box retailer and buy everything at one store. My wife shops at four different grocery stores because of the quality of the product and the prices of the product. We just are two different behaviors. Those are two different behaviors in the marketplace.

MR. SOLTANI: So the shopping behavior here would be, you know, building an architecture like I did to shop for a stapler, right?

To find different prices, you have to delete any of this information associated to you. And God forbid you use an authenticated source like Google or Facebook, that requires a login or is tied to your credit card, to your purchase device. So you have to then use anonymous cash and all this kind of stuff. I think that's a little much for most consumers.

MR. PRATT: Hmm.

MS. THOMAS: So I want to talk a little bit about Ashkan's presentation, which I think was actually incredibly -- not actually, it was incredibly helpful, in terms of understanding how the backend works, right? So thank you for that.

I think, you know, I agree with Stuart that loyalty is something that consumers respond extremely strongly to. They want to be, you know, in a loyalty program. They want to be given that additional offer because they shop in one place and not another on a regular basis, et cetera. And those programs have grown because that's what consumers have demanded, that if I'm going to be loyal to you as a brand, then you darn sure make -- make darn sure that I get something back for it.

Now again, we can talk about the outer limits of that when it goes beyond marketing, but as far as marketing is concerned, loyalty makes sure that you get better offers over time. But also let's remember, businesses want new customers as well. So a business is as likely to give a discount to a loyal customer, it's equally likely to give a discount to a first-time customer, to make sure that that customer gets involved and then will come back and become loyal in the future. So this isn't a matter of one or the other.

And just to sort of go back to the bottom line here, Ash made a really great point about the marketing offer. You might get one offer or the other, but then there is that firewall to whether or not you are actually going to be eligible.

And I think a helpful way to think about that is, you know, the Fair Credit Reporting Act, I'm not going to get into it, I promise, but it uses the term consumer-initiated, in terms of the kind of transactions that are covered by FCRA. The consumer initiates it and then it's not in the consumer's control what happens next, in terms of the eligibility decision.

In marketing, it's a marketer-initiated transaction, but it's the consumer who is in control of whether or not they respond to that marketing offer or they go around that marketing offer and say, that's not really what I want, I don't want 20 percent, I want 50 percent and I'm going to walk in the store and say so.

MS. ARMSTRONG: And Rachel, I think that's a really great point, but I think Pam and Ed are -- I would like to hear what they have to say, because I think they are talking a little bit about that very big fuzzy space in between the loyalty marketing and when you get into eligibility. So I'd like to --

MR. MIERZWINSKI: Well, sure. And you know, I think what Ashkan was talking about is not -- he's not against loyalty cards. He probably might have some, he might have a rewards credit card, I don't know.

But what he's against is consumers being selected based on secret profiles to be chosen to pay more. Nobody wants to pay more. Everybody wants to pay less. But nobody wants to be put in a box where they pay more. And in a nontransparent system, where thousands of bits of our lives are being collected about us, shared and used to decide who will pay more, my concern is not with big data per se, my concern is with, can we use big data in a positive way to promote financial opportunity.

I don't want banks to figure out which consumer we can ding for more overdraft fees. I want banks to figure out how to serve the under-banked, using big data to save money and encourage the use of the right accounts that make sense for that consumer to be able to build up assets.

Again, I have loyalty cards, I have rewards cards. I think that some seats on an airline, some seats at the last minute, seats right down in front at Yankee Stadium, are worth more than others. But I think a stapler is a stapler.

MS. DIXON: So this conversation has gotten to really the good part, really. So I'm just going to give the grab here. So look, big data is an opportunity for inclusion and it's an opportunity to help people. I've seen this with my own eyes, in other countries and in this country. It's incredibly important that we acknowledge that. And that when we target vulnerable populations or use sensitive factors that these are used with great transparency, oversight, and consumer control and are unfailingly beneficial to consumers.

So there should be no secret scores and there should be no secret factors. As a result, and to facilitate that, there's something that we can do, I think, that is a fairly -- it's a tweak. And I'm all for tweaks because they're doable, right?

So creators of consumer scores, whether they be static or femoral, enterprise, public, et cetera, if they stated the purpose of the score, if they stated the composition of the score and the intended uses of the score, and allowed uses of the score, then I think that that would go a very, very long way in transparency. And I believe that would also pull that under Section 5 of the FTC Act and then we would have some oversight.

We all know that Section 5 is broad and it would be very difficult to enforce a lot of this; however, it would provide a tweak and a first step toward bringing fairness while allowing benefits to occur. We have to have both, we have to have both. No secret scores and no secret factors.

MS. ARIAS: I think Rachel wants to respond, but I do want to followup on something that you just mentioned.

You said there should be transparency and consumers should know about it, but how do you communicate this to consumers? Given that there are so many scores and there are so many factors that go into these scores, and also that, you know, they don't even know they exist, right? So at what point do you communicate this information to consumers?

MS. DIXON: I think there's no perfect or beautiful answer to that. I think we're all struggling with that right now. How do you communicate in either a short or long or midterm or holographic privacy policy? Those things are, I think, in the midst of being decided right now.

But in the interim, a privacy policy would be a great place. This has its flaws, I am the very first person to admit this, but we need to start somewhere. And I'm all for starting with a tweak. Because protecting vulnerable consumers is a necessity, not an option. And as a result, let's start somewhere.

MS. THOMAS: So I just wanted to add, I couldn't agree more with Pam, the importance of making sure that when vulnerable populations are at stake or at target or the topic of conversation, transparency and making sure that they are not treated in a discriminatory way is incredibly important.

So I think the good news is that we do have, not just FCRA but many laws that make sure that that doesn't happen. There's FCRA, of course, the fair lending laws, of course, apply there. And I would argue that the Federal Trade Commission Act, the FTC Act, Section 5, if unfair or deceptive acts or practices -- if the marketing or the advertising is unfair to a vulnerable group in some way, that that would already be covered.

DMA also thinks this is incredibly important, such that we regulate that in our ethical code as well.

We have one of the first articles, in the almost 60 articles of requirements for any marketer doing anything, is disparagement. Make sure that any marketing that you undertake is not disparaging to any population and particularly vulnerable populations.

So this is something where, if there are problems, they are that we are not enforcing the laws that we have and that we should be in greater, you know, put greater resources toward that to make sure that the kinds of things that Pam is talking about are not possible. Protections exist, we have to make sure that they are acted upon.

MS. DIXON: I don't think most of these scores actually are protected, I really don't. I think they --

MS. THOMAS: Can you give an example of -- like, just for the sake of --

MS. DIXON: Aggregate. Aggregate credit scores, I think, are an excellent example and modeled credit scores. So they are used to provide offers of credit, and even to set initial insurance rates, right? But --

MS. THOMAS: So an offer of credit though could --

MS. DIXON: I'm not talking --

MR. PRATT: They're not used to set rates, they're not used to set rates.

MS. DIXON: They're not used to set rates for --

MS. THOMAS: -- would be covered by FCRA.

MR. PRATT: I just want to -- that's the bright line, I think, you're dealing with, right?

MS. DIXON: That's correct, that's exactly right.

MR. PRATT: Okay.

MS. THOMAS: Which is such an important bright line.

MR. PRATT: Correct.

MS. DIXON: We're at the razor edge of where the Fair Credit Reporting Act ends and something else begins. And I'm talking about the millimeter to the right where the something else begins. So I'm not talking about a firm offer of credit, I'm talking about the offers that really color and make a person's life different.

And we haven't talked a lot about health data. In the report we're coming out with next week, we talk a lot about how health data is being used in scoring. And this is health data that has been acquired outside of HIPAA.

And when health data is put into scoring factors, it is extremely prejudicial. And no one has -- there's not a law that covers this. Well, a little bit FCRA, but it's not applicable. So that's the piece that we're concerned about.

MS. THOMAS: I hear what you're saying, but I think, you know, I think the FTC has taken action in these areas, and very important actions, to make sure that FCRA is covering exactly where it needs to and that there is no gray area.

I think the Spokeo case was a really good example of that. You know, FCRA has worked for 40 years and I think that case showed that it continues to work. If third-party data is used for a permissible purpose under FCRA, then that data is a consumer report and the agency or the organization is a consumer reporting agency and FCRA covers it and the problem is solved.

So let's make sure that that -- you know, thank you. And let's make sure that those kinds of cases continue to happen where, you know, FCRA-covered data is being used in ways that are not in line with the law.

MS. ARIAS: I think Ashkan wants to jump in.

MR. SOLTANI: Yeah. So I'm --

MS. ARIAS: And Claudia.

MS. ARMSTRONG: And Joe.

MR. SOLTANI: So the backend verification happens through the FCRA-approved process, but if I'm never presented with the opportunity to apply for the lower interest rate card, kind of where in the -- how does that --

MS. THOMAS: There is no reason that you should ever have to take a marketer or anybody else up on the offer that you're given. That is completely separate from --

MR. SOLTANI: But if I don't see --

MS. THOMAS: -- your eligibility to get the product. So if you don't like the offer you have, that isn't the only offer that is available to you.

MR. SOLTANI: Actually, so --

MS. THOMAS: That's the law.

MR. SOLTANI: When I tested the kind of credit card sites, Chase for example, or CapitalOne, when you would repeatedly visit the site, you would never get, you know, one out of maybe 100 times, or statistically a low probability of times, you would get the other offers.

So as a consumer, if I reload the page, I'm not given the choice to say give me the zero interest card.

MS. THOMAS: Did you call the credit card company to ask what offers --

MR. SOLTANI: Do most -- I mean, so this rings to me a little bit --

MS. THOMAS: If I don't like the offer I'm getting, I go somewhere else to try to find an offer that I actually like.

MR. SOLTANI: Right, right. So let me kind of --

MS. THOMAS: That's what I'm getting at.

MR. SOLTANI: Let me give an analogy that might work. So everyone can vote, right? And you get verified at your polling place.

MS. THOMAS: Mm-hmm.

MR. SOLTANI: But you can make it very difficult for people to vote. I mean, we've historically seen people putting voting booths in places that are difficult for consumers to go to -- for voters to go to, it reduces the rate at which certain populations will vote, right?

And I feel like this kind of butts up against that. When you make it slightly difficult for people to do a particular outcome -- and sure, absolutely, you can call, you can scour the internet, you can research and find other offers, but we know most people don't call. Most people take the offers they're given and put some amount of credit in -- they put some amount of effort in their busy lives to do a set of activities, one of which is try to research a credit card and take the best deal that their search engine gives them.

MS. ARIAS: Why don't we have -- I think Joe hasn't spoken for a bit. Why don't we have Joe maybe jump in and then we can have Claudia.

MR. TUROW: I wanted to kind of pick up to where Ashkan was coming from. Putting it into broader historical perspective, I think we're really at a very different point in terms of how we understand pricing and what it all means.

I think if you look historically at the U.S. retailing situation, the 19th and 20th century was about the progressive, relatively speaking, democratization of prices. You could walk into a store and pretty well you would see the prices. Of course, some people got different prices, some people went into a back room. But generally speaking, prices became, for lots of interesting reasons, democratized.

We are moving away from that ideal in very interesting ways. And not just in the online or mobile space, but brick-and-mortars now change the prices by the hour and change the prices by the person. So literally, you could be walking through a store and the prices would be different for you, particularly with the new, you know, Apple and Bluetooth situation, where you walk through the store and it actually knows who you are. That's a very different way of thinking about the world.

And I think -- so there are some really, really important issues like financial and health, but in the broader sense of how we are going to see one another and understand ourselves. If we are walking through a world where we are consciously aware of, for reasons we have no idea, we are getting different offers, different deals, different understandings of us based upon calculations of our lifetime value, which is five years, by a retailer, that we have no idea why it came? That's a different mindset of how we understand the world.

And people are going to have to catchup with that. I think most of us are still in the 20th century and there may be good reasons why we're encouraged to be in the 20th century thinking about this. But the world is changing so drastically, it really creates incredible tensions.

MS. PERLICH: I actually really, really like that, because that's the direction I was thinking about as well.

I think there are two questions. One, are you comfortable with the brave new world of this kind of differential pricing? The question is, what role does alternative scoring play? Honestly, what Ashkan presented, it's not about alternative scores, it's about the big picture of using data. There isn't even a score. I mean, it's a flag on my user-agent. If you really want to make every single data point a score, then that would be infinite, you can't govern that.

He was really talking about the use of single data points. That's very different from the question of well-defined scores that are aggregates or even what I presented. My models would charge rich people more because they are more likely to buy at higher prices than poor people. So his findings are a human decision. There was an expert that thought, those guys must be rich, let's charge them more. There wasn't a model involved. It's the singular use of data points.

MR. SOLTANI: That's actually not true. So in the user-agents, sure, but in the credit card offers, remember X+1 was -- it determined, by some scoring, whether you -- your credit profile, for example. It would determine -- or the location, right? So the location would determine, based on some other factors, a score which is your distance to a competitor.

MS. PERLICH: I'm just saying it's a range.

MR. SOLTANI: For sure.

MS. PERLICH: It goes from a single data point all the way to these aggregate scores that --

MR. SOLTANI: Absolutely. Just clarifying.

MS. PERLICH: Yeah.

MS. ARMSTRONG: Joe, we talked earlier in some of our conversations -- oh, I'm sorry, Ed.

MR. MIERZWINSKI: Yeah. I just wanted to say that the issue -- I do care a great deal about the Fair Credit Reporting Act, but when we don't cross that line of determining eligibility or whatever, I care about the scores on the far side of the line.

I agree the FTC has done a good job with companies that have crossed the line, by selling information about your friends on Facebook or other social network sites that bears on your reputation and makes you a credit reporting agency. Good stuff.

But on the other side of the line, a lot of these credit card sites that you go to to find the best deal are actually what are called -- and I can't believe -- I don't think this term has been used yet today, lead-generation sites. And lead-generators auction you off, in realtime, not to the lowest or the best bidder for you, but to the highest bidder. And you have no idea, it's a completely nontransparent process.

Now sometimes they may send you to CapOne, in the credit card case, sometimes they may send you to a third-tier credit card company, but the lead-gen sites are also primarily used by the bad guys on the internet. The online payday lenders, the for-profit schools, and others.

And again, they are paying, based on all the information they collect about you, for a score that makes you someone that they can take advantage of. And that needs to be looked at. And fortunately, the states of New York and Illinois and other states, plus the CFPB and the FTC are looking into this. But it's a very important area, scores that are outside of the FCRA. When we get to a world where the FCRA is small and these other scores are big, that's a bad -- that's a bad world.

MS. DIXON: I think we're already at that world. And I want to go back to the point I made about the Klout score.

So in the report that we have forthcoming, we describe a gentleman who was denied -- actually, he lost a job offer and was told he was denied because his Klout score was deemed too low.

So it caused us to do a thought experiment, is Klout a CRA? And we had to come to the answer that it is not. Otherwise, every single entity on the planet would be a CRA. So we have to also apply First Amendment issues to this. There's a huge tension. We can't make everyone who uses a piece of data for eligibility reasons a CRA.

So given that, what do we do? And that's what I'm saying. And I think we really need to have transparency as an important first step. No secret factors, no secret scores. Tell people what's happening and then let's start figuring things out.

MS. ARMSTRONG: Thank you, Pam. Joe, I wanted you to take a little bit more time to speak about the future. I mean, how do you see some of these scores being used in the social, in the mobile, and other contexts?

MR. TUROW: Well, if we think about scores broadly, meaning indexes of how people act and predictive analytics, in terms of where they will go, my sense is that it simply has to become more and more pervasive. And the reason I say that is because of the hyper-competition that exists.

We're in a world today where the meeting between brick-and-mortar and the online world is so competitive. When you're competing with Amazon, that doesn't worry about margins, for example, seemingly, that raises lots of interesting questions.

And essentially I see it as mobile, wearables, and even cars. We're going to be in a situation where our car will be part of our decision-making and the decision-making about us. We'll be in situations where -- it is possible, if you give your permission, for cameras to look at your face. There is a company now, Emotion, that can actually look at facial features and decide certain aspects of what you think, your emotions, when you're purchasing something. There's a Russian company, and actually I think HP has a similar thing, which can look at you at check out and then try to connect you as you're moving around stores and elsewhere.

MS. ARIAS: Joe, can you lean into the microphone? We're having --

MR. TUROW: Oh, I'm sorry.

MS. ARIAS: -- trouble hearing you a little bit.

MR. TUROW: So the idea here is that, more and more, it is a question of not so much the technology, but what we want to put up with. The competition in the world is going to be that predictive analytics is the future. I think the 21st century is about data and data is incredibly important. Companies have to do this and they will do it as much as they can, with the idea of getting people to see that it's relevant to them.

And thereby -- and I won't go on and on, because it's a fascinating topic, but the issue of relevance is at the core. And it's terribly important, people want to get relevant ads. I think they want to get relevant offers and relevant deals. The tension that exists has to do with how do I do that, while not having my data being used for things I don't even know about and may not agree with. What is the seepage of those data elsewhere, okay? How am I being scored that may affect some other parts of my life? If I get a discount that's relevant, but I give in some data, well, that's terrific. I'll give you the stuff, but then it gets used in ways I don't want.

All of these things are part of the tension that we have to deal with in today's world. Everybody wants relevance, the question is do they really know the price of that relevance?

MS. ARIAS: And so I think that's an

excellent point, and I know we're running short on time, but we really want to talk about some of the solutions that folks think need to be implemented, to the extent we need any.

So I'm going to open it up to the panel, to the extent anybody has any thoughts on this. Ashkan?

MR. SOLTANI: So one idea might be, and this is, again, pie-in-the-sky and probably a few years out, if we can define the contours of the uses that we care about that are kind off limits. And we're talking about data and algorithms here.

You know, kind of like what I've put together, I wonder if there's ways to do basically unit tests? Test cases where you basically feed in data into an algorithm, feed in populations, feed in either fake users or real users or require auditing or reporting of the output, of the classification, such that you can audit the results and say, look, whatever the data input that you have, whatever those sources are, you're clustering these users based on race. Or you're providing offers based on sensitive categories of information that we don't want. And as a result, this algorithm seems to be discriminatory.

It's kind of pie-in-the-sky black-boxing, but it's essentially, I think -- because there are so many sources of data, because there are so many different algorithms, because often times the creators don't explicitly -- so the examples I showed were explicit. Often times, as Claudia described, it's clustering. It just so happens that you cluster and you're clustering based on distance from a competitor, but you also happen to cluster based on race or ethnic type, just because that also correlates to zip code.

And so if you start seeing that type of behavior in the output of an algorithm, then you can start saying, well, either the algorithm or the data sources are problematic. Sorry, that was a little --

MS. PERLICH: So this isn't necessarily an answer. I feel strongly that it really comes ultimately down to what you do. I think we have to focus on decisions that we are comfortable with making. I consider that there is the pie-in-the-sky, getting there is a long way from where we currently are.

The challenge of predictive modeling is it's a reflection of the current biases of human nature. It's a reflection of the fact that certain demographics are in worse economic state and the model will pick up on this and reflect that directly.

And I have a hard time making the model not do that. It's very hard for me to insert my morals into what just the state of the world is. My models come to the point where I am making decisions based on what that model tells me. I think that's, for me, where I can have a judgment easier, than trying to sift through the incoming and what it may or may not mean.

MS. ARMSTRONG: Everyone is going to have an opportunity, before we end, to make a final comment, but with what both Claudia and Ashkan have said, I wanted to raise an issue that the recent NCLC report described.

And it was a scenario where American Express lowered the credit limits because of other customers who shopped at places that consumers shopped and they had a poor repayment history. And this is also consistent with some of the other issues that that report raised about the discriminatory impact and I'd like to get some comments on that.

MS. DIXON: I'm concerned about cohort scoring, so who your friends are kind of tells other companies or entities or health care institutions who you are. It's a predictive modeling validation tool. However again, I come back to, we can't have secret factors and secret scores. Transparency is going to do a lot to ease some of this and I think we're going to have to find a way to meet in the middle.

You have to validate the score, right? But if you're validating a score in a discriminatory fashion, that's a huge problem. But we're never going to know about it unless someone has done some, you know, some very surgical-strike research.

MR. PRATT: So any transaction that AmEx would make relative to a portfolio of credit cards, we all know it, regulated by those same laws, that alphabet soup of laws we talked about before.

So there may be other instances that are outside of the credit portfolio context where I suppose you might have some conversation about, you know, the effects. But in the context of a portfolio, if you're going to change the contract, something about those terms, you have to control for all of the current law factors that are out there, ECOA, disparate impact, Fair Lending, anything else that would apply to that portfolio. So there's this penumbra of protection, if you will, around that kind of portfolio decision.

We don't know whether that was a good decision or not, we don't know if that was an effective decision or not, we don't know if that was an experimental idea in the midst of the recession, as every card issuer was trying to figure out what to do to measure risk, as card portfolios and, of course, all financial services portfolios began to crater just a little bit, you know, in terms of size of risk population, which was much, much larger than had been the case historically.

So I mean, there's a lot to think about. We're getting just this little tiny anecdotal moment here, but not much else really.

MS. ARMSTRONG: That's a good -- that's an excellent point.

MR. MIERZWINSKI: Well, I would just say briefly, that if companies do this and they do it on a non-protected class discriminatory manner, they may not implicate any of this alphabet soup of laws. Companies like LendUp and Move-In, if that's how you say that word, I've never used it before, are making decisions about who to make credit offers to based on their social networking status. But they're not credit reporting agencies, they're simply making decisions about their own customers, or potential customers, and they may not be regulated.

So we really need to look at regulating the system of scoring that isn't regulated today.

MS. ARIAS: And to follow-up on that, what do you envision would be the way to regulate? If, in fact, they are not regulated under the FCRA or any of the other alphabet soup laws?

MR. MIERZWINSKI: Well, I think the first step is really transparency of the online scoring system, the graphic that Ashkan put up of all of these hundreds of companies. They are all business-to-business companies, nobody knows who they are.

In the past, you knew that you had a relationship with your creditor. And although you didn't choose your credit bureau, you knew that you had the three credit bureaus and your creditors, who you could choose. But you didn't know about all of these other companies out there that were providing services. And that needs to be more transparent and consumers need to have rights when their information is used. The right to look at their profile of a data broker, the right to change the profile of a data broker, and the right to block the use of their information for other purposes.

And there should be some disclosure, just like there is when you are denied credit on the basis of a credit report, that you're denied credit on the basis of some other kind of report.

MS. ARMSTRONG: And with respect to transparency disclosure, I'm wondering about choice for consumers and whether you have an opinion about where in the ecosystem the consumer choice should occur.

MR. MIERZWINSKI: Well, we don't have a lot of time. If you're asking me, I would just say that I would refer people to look at the final Commission report on privacy. I think it's an excellent background around all of these questions.

MS. ARMSTRONG: Okay. I think it's time that we're going to run down this row and let everybody make a final comment before we conclude. So Rachel, you're first.

MS. THOMAS: Can we start at the other end this time?

MS. ARMSTRONG: Or would you like to be last? Okay, we can do that. Ashkan?

MR. SOLTANI: So one thing that I think that -- sorry. One idea that we might have -- I think Claudia is absolutely right that this is a difficult thing to get at and to underscore. One thing that we might want to do is to explore, like the FTC authority, for example, when a company says, you're getting the lowest price, right? And you're actually not, you're getting a different price based on ratings. Or Joe is getting a lower price than I am, should a company be required to say, you're getting the lowest price for you, based on this information? Or you're getting the best credit card deal for you, based on this information? Or can they just outright say, you're getting the lowest price, because that seems like an absolute statement.

And I wonder, under Section 5, if there is authority to at least nudge or poke at that.

MS. ARMSTRONG: Or you know, we can do this in any order, just as long as everybody gets to say something.

MS. PERLICH: Well, I'm just going to skip my right to have a final word here and leave it to the rest of the panel.

MS. ARMSTRONG: Thank you.

MR. TUROW: I'm just going to say something that may be quixotic, but we have found in survey after survey, I think five times, that people think that when a site has the word "privacy policy" on it, they think that it means the site doesn't sell or trade information without their permission. This is 56 to 62 percent of Americans. 75 percent don't know that, in fact, it's true that they give up -- that they don't do that, necessarily.

So the only thing I would suggest here, and it's only a beginning, but it would cause an interesting amount of chaos in the industry, which is if a site uses your information without explicitly getting your permission, it's deceptive by the FTC and it shouldn't be called a privacy policy. It should be called using your information. That would be an index to people to be careful. It would, right off the bat, say be careful.

If it says privacy policy, you know that you're safe. If it says something else, you know to be careful. And I think that would be the beginning of an interesting conversation with the industry.

MS. DIXON: So just to roll down the line, why not? So there are many new scores, many. Many, if not most of these, are outside of the current regulatory structure. Some of these scores are much more important than others. Some are not important, some are very important and the deciding factor is the impact. And we are focused really on eligibility uses, especially those that are happening outside of regulation, and they are.

We'd really like a solution, a first step of real transparency that's meaningful, no secret scores, no secret factors. Discussions with industries, so we know what's happening, and a real meaningful commitment from industry to have transparency about this.

We'd really like for there to be oversight and disclosure to the consumer about purposes, composition, and uses of the important scores. Probably not all of them, there's too many, but the really important ones that matter.

MR. MIERZWINSKI: Well, although I mentioned that I got a fishing catalog by mistake, I don't care about that kind of marketing, if I didn't make that clear. But I do think that financial information, healthcare information, and information about children needs to be looked at, in events like this, in greater detail. I'm encouraged by events like this being held. I'm encouraged that recently, it was in the press, that a number of civil rights groups have developed a platform on the use of big data for financial opportunity.

And we're starting to look at this, and we don't have the answers yet today, but it's encouraging that were starting to create a framework of answers.

MR. PRATT: So for me, as the CDIA, I'm just going to echo what I said at the very beginning, risk management matters. It's critical, it keeps us safe, and it ensures the transaction is me when it is me and it ensures the transaction stops when it's not me. So bank safety and soundness matters. It's important that we have ways to measure risk and to rank order consumers in terms of risk.

And just a couple of quick thoughts about the glossary, and I know this is kind of nerdy stuff here, but I hate the data broker term. It's a really sloppy term, it wraps around all different kinds of U.S. business models and it conflates issues in the public policy world as well. And we've seen this, actually, in legislation that's been introduced on the Hill. So it's just a crummy term. And I suppose if I could get out my big eraser, I'd erase the term data broker and we would try to parse through the issues in a little more refined way.

Get rid of the term score. The only reason I advocate for that is because it's too often conflated with credit score, too often conflated with what consumers now think of as a credit score. So again, this is just sort of marketing stuff. If we're communicating with consumers through venues like this, we ought to pick terms that make sense to consumers.

I think, Joe, your term about using my data versus privacy is the kind of example that pivots off of that same idea. Using my data is different than, you know, privacy, which may be a more amorphous term that we don't understand or connect with as much.

By the way, the problem with data brokers is it really doesn't -- you know, it kind of lumps together first party third party. We don't think those issues matter very much. Our members are third-party databases, dataflows have to come to third-party databases. It's how the American economy operates. Dataflows are necessary, I agree with what Joe said. Dataflows are going to occur, we're going to be in a highly competitive environment. I do think that what Joe said is right, consumers -- by the way, I'm going to add to what Joe said. I think consumers will catch up a bit, will begin to become smarter consumers, even in an environment where competition has ebbed, you know, changed a little bit over time.

So we should be encouraged. These dataflows can ultimately open doors for us that weren't open before, they can include us when we were excluded before, they can give us better offers that save us money. It's still up to us, as consumers, to do some shopping in the context of all that.

MS. THOMAS: So we've touched on some examples today, some of which, you know, at the end of the day, it sounds as though they are covered by FCRA, some maybe there are still questions about whether they're covered or they aren't.

But I think it's important to recognize that if we're focusing today on some difficult areas, there are a lot of areas that we are not concerned about, right? The dataflows are happening. Data is data is data. What we need to be concerned about is particular uses. Thank you for focusing us in that way today.

But I think it's also really important to recognize that the reason there are so few things that we have reason to be concerned about is because companies are making sure -- companies, nonprofits, all of the above, have strong incentives to self-regulate and not do bad things to their customers on a daily basis. And that there is DMAs, as well as many other self-regulatory environments, that are making sure that, if there are areas where the law doesn't cover, and there are so many laws that we discussed today that do, but in those areas, businesses are doing the right thing because they are being held to standards by organizations like DMA. And if they aren't meeting those standards, we know where to find the FTC and we do refer them over, as needed, to the FTC or other law enforcement.

At the end of the day, we need to enforce what we have, whether it's through self-regulation of the laws that we've got, and continue to make sure that data is used responsibly in both of those areas.

MS. ARMSTRONG: Thank you very much. And we'll continue down the line for our last few comments and then call it a morning.

First of all, I appreciate everybody coming. I wish we had more time, this was fascinating. We are accepting public comments about today's -- on these issues until April 19th. And thank you very much for everyone who sent a question card. I know we didn't get to everything, but we are thinking about these things.

And finally, I wanted to remind everybody about the next installment of our Spring Privacy Series on May 7th, which will focus more on health-type products.

MS. ARIAS: And just to finish everyone off, I really want us to pause for a second and thank this great, great panel.

MS. ARMSTRONG: Fabulous.

MS. ARIAS: Fabulous, that's right.

MS. ARMSTRONG: Fabulous.

MS. ARIAS: If we can applaud them, they did what we thought was unthinkable today, which is cover a really, really broad subject, cover lots and lots of issues, and really touch and started the conversation, so I really want to thank them.

But I want to thank you, our audience, and everyone out in the internet, watching us on the webcast, you've really been a fabulous audience and we really look forward to your comments and suggestions on this topic.

Thank you everyone.

(Whereupon, the proceedings concluded at 12:05 p.m.)

C E R T I F I C A T I O N   O F   R E P O R T E R

MATTER NUMBER: P145401
CASE TITLE: SPRING PRIVACY SERICS
DATE: MARCH 19, 2014

I HEREBY CERTIFY that the transcript contained herein is a full and accurate transcript of the notes taken by me at the hearing on the above cause before the FEDERAL TRADE COMMISSION to the best of my knowledge and belief.

DATED: MARCH 27, 2014

STEPHANIE GILLEY

C E R T I F I C A T I O N   O F   P R O O F R E A D E R

I HEREBY CERTIFY that I proofread the transcript for accuracy in spelling, hyphenation, punctuation and format.

SARA J. VANCE