FEDERAL TRADE COMMISSION

BIG DATA: A TOOL FOR INCLUSION OR EXCLUSION

Monday, September 15, 2014

9:00 a.m.

Federal Trade Commission

FTC Conference Center

601 New Jersey Avenue, N.W.

Washington, D.C.

Reported by: Jennifer Metcalf

FEDERAL TRADE COMMISSION

I N D E X

P R O C E E D I N G S

MS. GEORGE: Good morning. Good morning, everyone. It’s a few minutes after 9:00, so we’re going to go ahead and get started. Please take your seats.

Good morning, again. My name is Tiffany George and I am an attorney here at the Federal Trade Commission. Welcome to the FTC Workshop Big Data: A Tool for Inclusion or Exclusion. Before we get started I have a few housekeeping items to cover. Anyone who goes outside the building without an FTC badge we will be required to go through the magnetometer, an x-ray machine, prior to reentering into the building.

In the event of a fire or evacuation of the building please leave the building in an orderly fashion. Once outside of the building, you need to orient yourself to Constitution Center. Across from the FTC is the HUD building. Look to the right front sidewalk. That is our rallying point. Everyone will rally by floors. You need to check in with the person or persons accounting for everyone in the auditorium. In the event that it is safer to remain inside, you will be told where to go inside the building. If you spot suspicious activity, please alert security.

This event may be photographed, videographed, webcast or otherwise recorded. By participating in this event you are agreeing that your image and anything you say or submit may be posted indefinitely at ftc.gov or on one of the Commission's publicly available social media sites.

The Seasons Cafeteria is located inside this building and the operating hours are from 7:30 a.m. to 3:00 p.m. Please note that there are no food or beverages allowed inside the auditorium. Also, please remember to silence your devices.

And with that, now I'd like to introduce our FTC Chairwoman, Edith Ramirez, who will make some opening remarks.

(Applause.)

OPENING REMARKS

CHAIRWOMAN RAMIREZ: Thank you, Tiffany, and welcome everyone to our new facility, those of you who haven't been here before. I want to thank everyone for joining us here today for our workshop Big Data: A Tool for Inclusion or Exclusion. And I also want to take this opportunity to thank Tiffany George, as well as all of the other FTC staff members who have worked so hard to organize today's event, and also to thank the speakers for sharing their expertise with us.

We are at a pivotal stage in the Information Age. Thanks to smart phones and smart meters, wearable fitness devices, social media, connected cars and retail loyalty cards, each of us is generating data at an unprecedented rate. In 2013 it was reported that an astonishing 90 percent of the world's data was generated in the two preceding years. Today, the output of data is doubling every two years. Advances in computational and statistical methods mean that this mass of information can be examined to identify correlations, make predictions, draw inferences and glean new insights. This is big data. It has the capacity to save lives, improve education, enhance Government services, increase marketplace efficiency and boost economic productivity.

But the same analytic power that makes it easier to predict the outbreak of a virus, identify who is likely to suffer a heart attack, or improve the delivery of social services, also has the capacity to reinforce disadvantages faced by low-income and underserved communities. As businesses segment consumers to target what products are marketed to them, the prices they are charged, and the level of customer service they receive, the worry is that existing disparities will be exacerbated. Is this discrimination? In one sense, yes. By its nature that's what big data does in the commercial sphere. It analyzes vast amounts of information to differentiate among us at lightning speed through a complex and opaque process.

But is it unfair, biased or even illegal discrimination? And if so, can steps be taken to level the playing field? Those are the questions we'll be exploring today. Big data in its 21st Century form is in an early stage. We have the ability to shape its development and its outcomes. If we're alert to the risks presented by big data we can take steps to guard against them. We can help ensure that big data can be a tool for economic inclusion, not exclusion. That's the weighty subject before us today.

But before we begin the discussion, I'd like to address three questions. First, how did we get here? Second, what's our aim with today's program? And finally, where do we go from here?

Let me start by tackling the first question, how did we get here, very literally. Whatever mode of transportation you used to get to this workshop, there were apps or connected devices available to assist your commute. Those of you who came here using public transportation may have availed yourselves of apps to tell you when the next bus or train would arrive. If you came by car, you may have benefitted from GPS technologies that gave you directions, sent you realtime traffic alerts, or allowed you to summon a taxi or driver by tapping on a smart phone app. And for the virtuous among us who biked or walked here, you may have used a wearable device to track the distance traveled and calories burned. No matter your mode of transportation, once in the vicinity, an app or website may have helped you to find a spot nearby to buy a cup of coffee before arriving at the workshop.

These very devices and services that help many of us get here physically are also what brought us here figuratively. The popularity of smart phones and other mobile devices, the array of mobile apps we have at our fingertips, and the burgeoning internet of things phenomenon more generally means that countless individuals actively and passively generate information in an extensive ecosystem throughout the day.

The proliferation of connected devices, the plummeting cost of collecting, storing, and processing information, and the ability of data brokers and others to combine offline and online data means that companies can accumulate virtually unlimited amounts of consumer information and store it indefinitely. Using predictive analytics, they can learn a surprising amount about each of us from this data.

While powerful algorithms can unlock the value from immense data sets, their ability to draw correlations and make fine grain distinctions also raises the prospect of differential treatment of low-income and underserved populations. This is a risk suggested by the Commission's recent report on the data broker industry, the Commission's study of the cross section of nine data brokers, that data brokers aggregate online and offline data from disparate sources to make inferences about consumers' ethnicity, income, religion, age and health conditions among other characteristics.

As the FTC and others have found, some brokers create segments or clusters of consumers with high concentrations of minorities or low-income individuals. There may be legitimate reasons why businesses would want to sort consumers in this fashion, but the practice also raises the possibility that these segments will be used for what I've called discrimination by algorithm, or what others have called digital redlining.

We heard these concerns this past spring at the FTC seminar on predictive scoring. There are now products beyond traditional credit scores that purport to predict or score everything from the chances that a transaction will result in fraud to the efficacy of sending consumers catalogs and the best prices to offer consumers. Some speakers lauded the benefits of such predictions, emphasizing that they enable the personalization many consumers want and help minimize the risk of fraud. But other speakers worried that certain predictive scoring products could fall outside the reach of the Fair Credit Reporting Act and the Equal Credit Opportunity Act, despite having an impact on consumers' access to credit, housing, employment and insurance.

For example, if a company lowers my credit limit based on a score that reflects my own credit history, I would be entitled to certain protections under the FCRA. If, however, the same company lowers my credit limit based on the scores of a group of which I am a member, the application of the FCRA may be less clear. Will these scores be used in ways that influence the opportunities of low-income, minority, or other populations to get credit, jobs, housing or insurance in ways that fall outside of the protections of the FCRA or ECOA? Could the use of geographic information, such as zip codes, for example, lead to Americans in low-income or rural neighborhoods being charged higher prices? And if so, is this a worrisome function of big data or a just a continuation of age-old pricing practices and market forces?

These and other issues figured prominently also in the White House's wide-ranging report on big data, which squarely raised the concern that large-scale information analytics will be used for disparate or discriminatory outcomes for certain consumers, even absent discriminatory intent. It's these questions and concerns raised by these prior initiatives that bring us to today's program and to my second question, what is our goal today?

We'll explore whether and how big data helps to include or exclude certain consumers from full opportunity in the marketplace. And to help shed light on these issue we've convened experts from industry, consumer, and civil rights groups, academia and government, all of whom are representing a wide variety of perspectives. Our panelists and speakers will provide us a framework for our conversation today, assess current big data practices in the private sector, discuss possible developments on the horizon, present pertinent research and offer potential ways to ensure that big data is a force for economic inclusion. It's my hope that our participants will discuss in depth the benefits and risks of big data to low-income and underserved populations.

On the benefits side, let me start the discussion with one example. New York City is developing a tool that combines eviction data with emergency shelter admission information and other data to predict when individuals or families are on the brink of homelessness.

Using this information, the city is able to deploy social workers to help these families and prevent them from ending up on the street. This is an example of positive government use, rather than a business use, but I hope our speakers -- our speakers will provide examples showing how companies can also use big data to benefit those in low-income or underserved groups.

And as for real world examples of possible risks, let me cite a study conducted by Latanya Sweeney, who's here from Harvard serving as the Commission's Chief Technologist. Professor Sweeney found that web searches for distinctively black names were 25 percent more likely to produce an ad suggesting the person had an arrest record, regardless of whether that person had actually been arrested, than web searches for distinctively white names.

This could have devastating consequences for job applicants and others by creating the impression the individual has been arrested. While the research did not establish why the algorithm yielded these racially disparate results, it does provide a concrete example of how an algorithm may have adverse repercussions for a particular population. I expect we'll hear more illustrations today, including from Professor Sweeney who will be presenting results of a more recent study.

After we conclude our workshop, the question naturally arises, where do we go from here? We may all have an array of apps to guide us home when we leave this afternoon, but there’s no clear path for navigating the use of big data in a way that advances opportunities for all consumers while diminishing the risks of adverse differential impact on vulnerable populations.

We may not yet know what the best course ought to be, but I believe we should have at least three objectives going forward. First, we should identify areas where big data practices might violate existing law. Where they do, the FTC is committed to vigorous enforcement of the law as demonstrated by cases such as our recent action against Instant Checkmate, a website that promoted some of its background checks as tools for screening tenants and employees. The FTC alleged that Instant Checkmate did so without regard for the FCRA, and we obtained a $525,000 fine and a permanent injunction against the company. In addition to helping the FTC and others to enforce existing laws, today’s program should also help identify any gaps in current law and ways to fill them.

Second, we need to build awareness of the potential for big data practices to have a detrimental impact on low-income and underserved populations. I’d like today’s program to help foster a discussion about industry’s ethical obligations as stewards of information detailing nearly every facet of consumers’ lives.

Third, and relatedly, we should encourage businesses to guard against bias or disparate impact on low-income and vulnerable populations when designing their analytic systems, algorithms, and predictive products. A good example is the Boston Street Bump App highlighted in the White House Big Data Report. Like any big city, Boston has its share of potholes and faces the ongoing challenge of staying on top of street repairs. To help address the issue, the city released a mobile app residents could use to identify potholes in need of repair.

But the city also recognized that because lower income individuals are less likely to carry smart phones, the data might skew road services to higher income neighborhoods. They addressed this problem by issuing the app to road inspectors who service all parts of the cities equally and used the data gathered from the inspectors to supplement what they received from the public. This illustrates how considerations of risks before launching a product or service can help avoid them.

So, big data can have big consequences. Those consequences can be either enormously beneficial to individuals in society or deeply detrimental. It will almost certainly be a mixture of the two, but it's the responsibility of the FTC and others to help ensure that we maximize the power of big data for its capacity for good while identifying and minimizing the risks it represents. As we navigate the transformative terrain of big data, it's vital that we work to ensure that technological innovation benefits all consumers whatever their backgrounds.

I look forward to hearing the thoughts and ideas of our panelists on how to do just that. And I thank you all for your contributions to that endeavor. Thank you.

(Applause.)

CHAIRWOMAN RAMIREZ: Let me hand it back to Tiffany.

MS. GEORGE: Thank you, Chairwoman.

We'll now begin with our first presentation, Framing the Conversation, which will be lead by Solon Barocas, a Post Doctoral Research Associate at the Princeton University Center for Information Technology Policy.

PRESENTATION: FRAMING THE CONVERSATION

MR. BAROCAS: Good morning. Let me begin by saying how thankful I am to be here. I really appreciate the opportunity to speak with you all. And I particularly want to thank Katherine and Tiffany for putting together what I think will be an excellent day. I am Solon Barocas. I'm a post-doctoral fellow at the Center for Information Technology Policy at Princeton, and I will be presenting today what I hope will be a way of framing the conversation today and hopefully going forward as well. This draws on some of the work that I've been doing, and I encourage people who are interested in what I'm presenting to take a look at my website where you can find this paper if you want to follow along while I present in more detail.

But let me begin. Okay. So, big data -- we've come, I think, to know the three Vs as a common definition. That the volume of data is exploding, that the velocity at which the data is accumulated is increasing, and the variety of formats of data is also likewise proliferating. This is a useful definition, but I tend, I think, to focus instead on the traditional categories from the social sciences, observational data, what we might call self-reported or user-generated data, and experimental data.

And what I mean by this, then, is that there are actually three valid, different things happening here, all of which have interesting consequences for consumer protection. One is that there are many more ways to actually observe consumers and consumer behavior, things like transactional data, but of course, we can now think of things like mobile phone and various health devices, self-reported and user-generated data being the vast variety of social media that people use. And finally, experimental, which I think has now become slightly more familiar to people in the wake of this Facebook experiment that got a fair amount of press. And what I mean by that is there are now platforms upon which to perform large-scale experiments in the wild in ways that were basically impossible before. And I think these are the useful ways, perhaps, to think about it.

For our purposes today, however, I'm going to focus on data mining, this is the more traditional term from industry and the academy, which is in some ways what we might call a subfield to machine learning, which is a -- a kind of field within computer science that is devoted to the automated computational analysis of large data sets. And again, I focus on this, in part, because I think for our purposes today it is the analysis and use of the data that is interesting, perhaps less so the technical challenges that large data sets present to those who accumulate them. So, the remainder of my talk will focus specifically on the analytic techniques and why those analytic techniques present some kinds of trouble for us when thinking about consumer protection.

So, what I'll say then as a kind of starting place that we can define data mining as the automated process of extracting useful patterns from large data sets, and in particular, patterns that can serve as a basis for subsequent decision making. You can -- I'm saying here in quotes "learning," meaning I learned from the previous examples that there is some general trend, some relationship in the data that I imagine will hold true in the future and I can use that as a way to make future guesses and inferences as mentioned earlier already.

For terminology, I thought I'd also point out that within the field this accumulated set of relationships within the data is commonly referred to as a model. So, you might have heard the term predictive model. What that refers to, then, is all the various kinds of patterns that have been extracted from the large data set that then inform future decision making. And this model can be used in a variety of ways.

To begin with it can be used to classify entities. So, the most common example of this would be spam. I think many people are familiar with this. Your computer often, webmail in fact, will make guesses about whether or not your message is spam or not, and again, it arrives at a rule to determine what is spam and what is not spam based on the history of examples of spam it has. Likewise, it can estimate values of unobserved attributes, or it can guess your income, for instance, as also mentioned. And finally, it can also make predictions about what you're likely to do. So, future consumer behavior of all sorts.

Now, you might say, as again was already mentioned, that, of course, data mining is discriminatory. The very intent and purpose of the activity is to be able to differentiate and draw distinctions. And what I would say, too, is that it is in some sense a statistical form of discrimination that is almost by necessity a rational form because it is being driven by apparent statistical relationships. And the data -- these are not arbitrary or this is not a case of caprice; this is, in fact, evidence suggesting that there are reliable patterns to the data. And using that you can confer to the individual those qualities which happen to be similar to those who appear statistically similar. So, if I reside in one particular statistical category that has been revealed by the analysis, they can impute to me those same qualities.

So, the remainder of the talk will focus on this five-part taxonomy, which is me basically trying to explain how the process of actually mining data lends itself to a variety of issues that can raise concerns with discrimination and fairness. So, let me jump right into it.

Again, a technical term is "target variable." What this basically refers to is when I set about trying to determine if there are useful patterns that correlate with some outcome, I need to be very specific about what I mean by the outcome. So, when I am looking for good customers, I actually need to arrive at a formal definition of what good customer means. Does good customer mean that it is the one from whom I can extract the most profit? Is it the one I can have a long-term relationship with? Is it the one that if I provide some inducement will stay a customer? And there's no way to actually avoid this formalization process. You must specify in some definable way what it is that you are looking for. And so the exercise of mining data always begins with actually having to establish some translation from a business problem into a problem that can be solved by predicting the value of this target variable.

And in general, the art of data mining -- the kind of creative work of data mining involves this process of translation, finding a smart, clever way of actually translating some kind of business problem into one that can be solved by predicting the target variable, by inferring the value of the target variable. And I think here's what's interesting is that the way that the business goes about defining the target variable can have serious consequences for whether or not the data mining process has a disparate impact.

In my own work I look at employment, and you might say that trying to predict whether or not someone is going to be particularly productive as compared to predicting whether or not that we're going to remain a customer -- rather, an employee for a set period of time, trying to avoid turnover, for instance. Those differences and definitions will have very different consequences for how you rank potential applicants. And the same would likewise be true with consumers.

The second part of the taxonomy is what, again, data miners refer to as training data. Training data is the large set of information that you use to extract some kind of useful rule. It is the set of examples that you look at in order to decide if there are actually useful patterns to guide future behavior, future decision making. And I think, in this case, there are really two different, although related, problems with training data that again can have consequences for fairness. One is that, as also mentioned, that the -- the set of examples can be skewed in some way. And the second, that the examples that you draw on could actually be in some way tainted by a prior prejudice.

So, let me try to walk through this a bit. When trying to derive some general rule from a set of particular examples, the only way that rule will actually generalize to future cases is if the particular set of examples happens to be representative of future cases. And as we know from Latanya Sweeney's work, this main -- rather, from the Street Bump case, we know that this is not always the case. And, even more interestingly I think, often times companies are in the position of -- are often seeking ways to try to change the composition of their customer base such that to suggest that you can draw general rules from what customer base that you are purposefully changing, again, to put into doubt the idea that this is representative data; that, in fact, you're dealing with a subset of all possible customers, and the particular subset you're dealing with changes over time.

We could also point out, I think, that the reason why data is unlikely to be particularly representative in certain cases, that is for reasons having to do with the following. So, to begin with, it might well be that certain populations are less involved in the formal economy and their various mechanisms in producing these kinds of digital traces. They might have unequal access to -- and less fluency in the technology that's required to produce those kinds of digital traces. And finally, they simply might be less profitable or in poor constituencies and, therefore, not the subject of ongoing observation.

And I think that the serious problem here is that often times the under or over representation of particular populations is not always evident. Sometimes when a geographic distribution is skewed in some obvious way, as in Street Bump, we might have intuitions that, in fact, there is a problem, but many times it will be much, much more difficult.

Finally, you could also say, then, that when you have this skewed example, it also suggests that companies should be devoting their attention to some populations and not others. And over time this can have a compounding effect where certain populations are discounted further and further because you have less and less opportunities for those populations to disprove your sense that they are not, in fact, good customers. You are in fact, limiting the opportunities to buck the apparent trend. And this is a serious problem in credit scoring where the industry has long worked on problems trying to deal with that.

Labeling examples. This is the process of actually trying to specify what is, in fact, a good customer and what is, in fact, a bad customer from examples. So, I mentioned the example of spam. Let me actually jump to this example. So, during the debates leading up to the Equal Credit -- Equal Opportunity -- no -- Equal Credit Opportunity Act, Fair Isaac pointed out in those congressional debates that in fact any way of drawing some rule about how to extend credit to customers that looked to previous ways that consumers were evaluated as potential customers of credit would simply reproduce any prejudice involved in those past decisions, meaning Fair Isaac could not simply draw on the history of credit decisions to automate the process; it actually had to find new ways to decide what, in fact, is a good target for credit. And what this reveals, then, is that any decision that uses past uses as a basis for inferring rules must be sensitive to the fact that those decisions might be tainted by prejudice in some way.

Finally, in this same theme, along the same line, we can point out then that it’s not only the case that data mining can inherit past prejudice, but it can continue to reflect the persistence of prejudice in the behavior, taken its input to some kind of model, and this, I think, is a way of categorizing some of the work that Latanya Sweeney and others have done showing then that if the input the algorithm receives is itself biased or prejudiced in some way it will simply be reflected back in the recommendations of that system.

Feature selection. This is the process of deciding what variables, what criteria associated with each person will you actually fold into your analysis. And here again, I think this is an interesting issue because you would imagine that big data presents opportunities to vastly increase the amount of features and variables you consider. Of course, these -- of course, the addition of the -- adding additional features to the analysis can often be costly.

And it may well be that your analysis does very well when considering a certain set of features, but it doesn't do particularly well for some populations because it doesn't actually carve out the population in a particularly precise way. Redlining is the traditional example of this. Using neighborhood alone as a way to decide who is worthy of credit is an extremely coarse way of making that determination. And I think those same kinds of problems can actually translate to this new area because it is still possible that additional data would be useful in drawing distinctions for particularly marginalized populations that simply might just be very costly. It might be very difficult to obtain that information. And the question therefore becomes, I think, does it justify subjecting these populations to less accurate determinations simply because it actually costs additional money or resources to gain that kind of information?

This fourth point of the taxonomy is what we call proxies. And what this refers to is the fact that often times many of the features that are legitimately relevant in making some kind of predictions about customers might also be highly correlated with their class membership, meaning certain features, certain attributes, are both proxies for the thing you care about and proxies for the person’s class membership.

And what’s worrisome here, then, is that it may well be this is actually simply reflecting the fact of inequality in society, and it’s a particular form of inequality where members of historically marginalized and protected classes are disproportionately in a less favorable position. And big data is in the position potentially to simply further expose the exact extent of that inequality.

I will, in the interest of time, jump over this. The final part of taxonomy is masking, which refers to the idea that it is possible to mask intentional discrimination by relying on any of the number of ways I've identified here of having discrimination happen unintentionally. Decision makers additionally can rely on data mining to infer whether or not you belong to a protected class and then to use that information in secret to discriminate against you.

I want to emphasize, though, and this is I think one of the most important points I'll make today, is that unintentional discrimination of this sort identified in the first four parts of the taxonomy is far more likely to be occurring, and it has potentially far more consequences than the kinds of intentional discrimination that could be pursued through masking.

And I'll simply conclude by saying that I think there's a serious issue here about the unintentionality of the discrimination that might be occurring. And in my own research I have looked at Title VII and in employment decisions, and my sense actually is that this aspect of the problem, the unintentionality of the problem will pose serious issues for trying to bring to bear legal remedies. It's unclear that we have the tools when looking at existing laws to actually address this form of unintentional discrimination.

Additionally, if the problem is that we are exacerbating inequality, it's also unclear whether or not using discrimination laws as a way to deal with that issue is the correct mechanism.

And finally, I think for many of the kinds of problems identified earlier there's no ready answer, both at a technical and, I think, legal level, and we really require, I think, a conversation that involves both parts of this debate, the technical and the legal dimension.

So, thank you very much, and I hope people will speak with me if they have further questions. Thanks.

(Applause.)

PANEL 1: ASSESSING THE CURRENT ENVIRONMENT

MS. ARMSTRONG: Welcome, everyone. I'm Katherine Armstrong from the Division of Privacy and Identity Protection, and I have to say we've been looking forward to today for a very long time. And so, thank you all very much for coming and welcome to Panel 1.

Today we -- this panel is going to examine the current uses of big data in a variety of contexts, from marketing, to credit, to employment, and insurance, and how these uses impact consumers. Today we hope to do one of the things I think the Commission does best, and that's to ask questions, to listen, and to learn. Before I introduce the panel, I want to remind everybody that

Solon’s PowerPoint or his slides are available on our website and well worth studying, as well as his paper.

So, let me briefly introduce our panel, and then we’ll begin. Kristin Amerling is the Chief Investigative Counsel and Director of Oversight for the U.S. Senate Committee on Commerce, Science, and Transportation. danah boyd is a Principal Researcher at Microsoft Research and a Research Assistant Professor at New York University. Mallory Duncan is the Senior Vice President and General Counsel at the National Retail Federation. Gene Gsell is Senior Vice President for U.S. Retail and Consumer Packaged Goods at SAS. David Robinson is a Principal at Robinson + Yu. And Joseph Turow is a Professor at the Annenberg School for Communication at the University of Pennsylvania. So, welcome and thank you again for agreeing to participate in this panel.

I’m going to start with a question about what is big data. What makes this data unique? Is it the three V’s, velocity, variety and volume, or does it have something else to do with the relationship derived from making connections among data sets? And you’re all free to speak to that, or whoever wants to jump in first.

MS. BOYD: Yes, I’ll jump in. I’ve been -- so, I have a mixed background. I started out as a computer scientist, I retrained as an anthropologist. So, I look at big data from both of those lenses. And we can look at the technical phenomenon, and much of what Solon referred to gets at that, but there's also a social phenomenon, which is, in many ways, tethered to the hopes and dreams and fears and anxieties associated with big data. All right.

The possibility that we will get to a perfected idea of statistical knowledge, that this will give us a new form of fact that will allow us to make meaning of the world around us, which, in many ways, obscures the complexity of probabilistic information -- right -- which is a lot of what we're dealing with is probabilistic. The data is imperfect, you know, just like Solon was talking about.

And so for this reason, I like to think of big data not simply in its technical sensibilities, but as a socio-technical phenomenon that brings with it a lot of different confusion and chaos. I bring this up because I think it's really important to remember this, especially in light of the conversation we're having today, because a lot of what goes on is the uncertainty, not necessarily the formalistic mechanisms of data mining, data collection, or data analytics.

MR. ROBINSON: And if maybe I could just briefly pick up on that, I think one of the things that

Solon mentioned that I think is extremely important that was also central to the FTC's Report is that in some of these cases you have data that was gathered for -- initially for some purpose that didn't require high fidelity, like slightly making more accurate the list of people that you send out a mailer to. And now, in some instances, some of that data is being used for purposes, like, deciding that certain people are likely to be fraudsters and will not be transacted with by actors in the marketplace.

And I think one of the great concerns that the civil rights community has is to make sure that where we're confident -- well, I'll speak only for myself -- I'm confident that businesses are going to do things in ways that are optimal from a financial perspective, that if something helps to make something more profitable, that it will happen. But I think, you know, what is the harm from a civil rights perspective versus from a business perspective when the occasional minority or unbanked, or underbanked, or otherwise marginalized person is incorrectly excluded from some product that they'd be ready to transact with. You know, at some level some amount of that is a cost of doing business. And I think one question is whether the amount of that that's acceptable as a cost of doing business is the same or is different than the amount that is acceptable as a civil rights' matter.

And I'll just say -- I mean, we -- our group of technologists that works with civil rights folks released on Friday a new report on big data and civil rights, which you can find at bigdata.fairness.io, which does our very best to sort of inventory these concerns.

MR. GSELL: So, I'd like to go back for a second to what is big data? Data's been around for a really, really long time. And people have been using it and analyzing it and trying to figure out what it means and what they should do with it.

Today, there's just more of it. This phenomena that this new thing called big data has existed, it's not something that just came into vogue; it's something that's been around a long time. And big data, by real definition, is more data than your organization can handle. Okay. I mean, that's big data. So, if you've got more stuff coming to you at home than you can deal with, you have big data.

The question really becomes, as more and more data sources become available, more and more data is out there, how do you gather it and make sense of it? I think the -- I think an awful lot of people give the industry more credit for sophistication than actually exists. Most people for the most part are still somewhat overwhelmed and a bit behind the curve on the notion of dealing with all of the new informational data that's coming through.

MR. TUROW: Can I just pick up on that? I agree, and I've talked to a lot of people who say exactly what you say in the retail business; for example, that they're overwhelmed and that we're at baby steps now. But it's the beginning of an era. And I would object to the notion that big data are simply the continuation in volume, because when you start adding velocity, and volume, and variety, and the notion then becomes predictive analytics, we're in a different world.

We're in a world where hundreds and hundreds of data points are used to come up with conclusions about people that are almost not even intuitive a large part of the time. You come up with the -- you have a key indicator that you're trying to look for, but the notion of which data are going to be used in the end -- an example, which may sound crazy, but I -- you know, it's not totally nuts.

Let's say you're a retail establishment, and you're interested in trying to predict which people are going to become less-valued customers, and you have a definition of a less-valued customer. You run your data with your hundreds of thousands of customers and you find that people who start buying vegetable seeds for planting in an urban environment predict that they are going to become less-valued customers in the sense of giving back more stuff, you getting only for sales.

Now you might say, what does one have to do with another? I could think -- and this gets back to what danah was saying, there are lots and lots of reasons we could think about, and I could give you some, as to why a person buying vegetable seeds would be predictive as a customer you wouldn't want to deal with the way you deal with other customers, giving discounts and other things like that.

But from the big data standpoint, the key is it's predictive, okay. We may not be sure why it's predictive and it gets used like that. And the notion of personalizing data that way is a terrific change in the way companies begin to evaluate their customers on many different levels.

MR. DUNCAN: Let me just say a couple words about the retail industry. Obviously, we operate on a very narrow profit margin. It's about two percent on average. And so it's important for the industry that we're able to find those customers who are going to be long, loyal, valuable customers.

When we talk about big data, in a sense, we're really talking about an expansion of what's always been done in the retail industry. If you go back a hundred years and you think about how your typical store worked, the store manager was constantly analyzing the shoppers in his store and trying to determine what is it I have to move in the store in order to attract more people; what is it I have to say to this customer in order to increase the loyalty. What big data, or what's referred to as big data, is an expansion of that effort. They are new analytic tools in order to accomplish the same thing. If we're not able to bring people in the store and not able to get them to increase what they're spending, then chances are the store's not going to survive.

MS. BOYD: I think this actually raises a different question which is tethered to the topic of today, which is, how do we evenstart to measure or make sense of fairness? Which is usually where we're starting to think about sort of the challenges of how big data gets used.

Now, in the American historical context we usually have a battle between equality and equity as our models of fairness, right? Equality is the idea of equal opportunity, we create that even playing field, everybody enters the table at the same fair starting point, and that's how we constitute fairness is when we have equal opportunity. Equity, of course, is saying, guess what, we have a large amount of systemic issues that result in the fact that people do not enter the table at the same playing field, or same level, so how then do we think about offsetting or dealing with those structural issues and how do we think about reconstituting, you know, the societal infrastructure so we can think about fairness, right? And mind you, we have a long debate in the U.S. on this issue of equity. Right. We get into this discussion of affirmative action. We get into this discussion of whether or not that constitutes socialism, and politics, politics, politics.

But there's a third logic that big data brings to bear with what we talk about as fairness. Something that is very much coming from the market-driven logic that Mallory talked about -- right -- which is the idea that we're trying to optimize out efficiencies, and to think about distribution of limited amounts of resources. Think about how we allocate in the best way possible in order to either maximize profit, minimize, you know, law enforcement officers on the street; you know, in another context, thinking about how we distribute resources or maximize opportunities.

The challenge with that is that market-driven logic of fairness often really comes up pretty viciously against our notion of what is equity, because of the fact that, as Mallory pointed out, we have these really small margins. And the question, then, is who bears the responsibility for, you know, the fact that we have, you know, retailers who need to figure out how to be profitable? I mean, we have the fact that many, you know, of our customers are not going to be that profitable element.

We've had this historically, right? Where do we actually allocate new, you know, stores? Do we do it in a way that is near neighborhoods who are not considered profitable? How, then, do we think about the social ecosystem? The reason I bring this up is because big data is , when used well, when the predictive analytics are done right, when the data mining is done with some level of statistical accuracy, you can get to a point of all of that unintended discriminatory or unfair outcomes because of the fact that we're trying to minimize -- you know, you're trying to maximize profit, minimize, you know, risk, and really deal with those efficiencies. And that's part of the trade-off in a commercial setting.

MS. ARMSTRONG: And we're going to be following up and circling back to the fairness and ethics as we continue on with this panel, but I think that's an important issue to bear in mind because it resonates through all that we're talking about.

I'd like to ask Kristin to also describe a little bit some of the findings of the Senate's Big Data Report last year.

MS. AMERLING: Sure. I'd be glad to, and thank you for the opportunity to participate today.

Chairman Rockefeller, as Chair of the Senate Commerce Committee, recently conducted an inquiry into how consumer information is collected, analyzed, shared and sold that I think shares the goal of this panel today, which is assessing what is the current landscape here. And just to give you a little bit of background, the inquiry was conducted by reaching out to nine major data brokers to ask what are their practices in obtaining, analyzing and sharing consumer information. And Chairman Rockefeller released findings in a report at the end of last year, a majority staff report.

I think that there are four major findings that are particularly relevant to the discussion that we're having on this panel and today.

First, companies, data brokers that collect information without direct interaction with consumers, and often without their knowledge, are collecting a tremendous volume of data and it has tremendous specificity.

Second, the companies are collecting this information from a very wide variety of sources.

Third, the result of analyzing this information that is collected includes products that are lists of consumers that define them by characteristics that include their financial and health status, including groupings of consumers based on financial vulnerability and other vulnerabilities, and they include another set of products that the Chairwoman referred to this morning relating to scoring consumers, predicting their behaviors based on data that's collected. And some of these products very closely resemble credit scoring tools that are regulated by FCRA raising questions about how these products that may or may not fall under the FCRA are being used.

And finally, the fourth finding that I think is worth noting is the lack of transparency that consumers have into data broker practices. And I'm happy to elaborate a little bit more on the four points.

MS. ARMSTRONG: Well, you know what, why don't you weave them in as we continue the -- the conversation?

MS. AMERLING: Okay. Sure.

MS. ARMSTRONG: But raising one of the points that Kristin just brought up, I wanted to also throw out to the group whether where -- whether where the data comes from matters? Whether it's coming from internal sources, external sources, third parties, whether it's passively collected or actively collected? Does it matter in terms of use or types of information?

Joe?

MR. TUROW: Yeah, I think it matters a lot, but I think we have to be careful to say that just because a store, for example, collects the data, it's not a problem. The example I gave with the seeds -- just to push that a little bit forward -- could reflect a hidden discrimination.

Let's say a person begins to plant a garden in her urban area because she's just lost her job, has to take care of her grandchildren. Those kinds of subjects can be brought out, not in direct discrimination, we know this person has lost her job, we know this person had to take care of her grandchildren, she has no husband or whatever, but rather, the fact that she's buying vegetable seeds. You see, it's the idea of hidden discrimination even within a particular store.

Now, add to that the things that you can buy from third parties that could build even greater profiles about people without anyone knowing that it takes place.

People going through stores with loyalty cards, and then the material gets put on top of that which can lead to many types of discrimination that we have no clue about.

MR. GSELL: So, that's certainly a possibility, I mean, the inherent when you do analytics on data, but one of the things that really is driving a lot of the change is the ability to process all of this data. It's one thing to collect it; it's another thing to actually do something with it, okay, and I would contend that the ability to tease out -- actually, to eliminate the need to sample. So, historically, data was so big that you did samples, and inherent in samples are some of the biases because they're based on how the sampler decides to set up their sample set.

When you have big data and you have the ability to use what I'll call "big compute against big data," you eliminate the need for sampling. And when you eliminate the need for sampling and you go against the entire data set, you have a much greater chance of eliminating historic bias that have existed based on the way people have decided that this represents an entire population. You don't have to represent an entire population anymore. With big data and big analytics, you can hit the whole thing.

MR. TUROW: But that's my point. See, that's exactly what I'm saying. What I'm saying is that increasingly companies -- and now it's harder, five years from now it will be easier -- companies will be able to use data in variety, velocity, and volume in such a way as to personalize a model. So that if I find that there are a thousand characteristics that I can bring together and come up with just a couple that make me decide that I should go after you, that may be a discriminatory decision and you don't even know it, because the -- the data that you're using are so part of the person's life in secondary ways that they discriminate even though it's not said that it's a low-income person or a person of a certain minority group. It just shows up that way.

MR. GSELL: I think you'll give us more credibility or ability than actually exists.

MR. TUROW: Okay. The last thing I'll say about that is you're right, but what's happening is, what is the trajectory of interest? And if you look at what people in the business are saying, that's where they want to go. They don't say they want to discriminate, but they want to say we want to be able to predict what a person is going to do when that person is walking into a store.

Eric Schmidt at one point said about Google, we want you to go to Google to find out what your job should be in the future. Okay. That’s what he said several years. We want you to go to Google to find out what your career ought to be. That’s quite a statement. They can’t do it now.

MS. BOYD: So, one of the things that’s important to understand is that the data that we’re talking about is not just about the data that you may give to a company or a data broker or even your interaction purely with them, but in many ways it’s about how you fit within a network of other actors and what else they’re doing, right?

Historically, we understood this is categories and, in fact, a lot of our conversation about discrimination is a conversation of how one fits into a protected class or a protected category, right? And you think about categories as a way of bucketing. And this had to do with the fact that we didn’t have the whole data set and that we couldn’t actually imagine the kinds of personalization that we’re talking about.

Personalization is only made possible because you actually can position somebody in relation statistically through a whole variety of other actors through networks, networks that in many ways are not intentionally designed for by the system creator. They’re looking, literally, for correlations that they can see are probabilistic connections. But this also means that we're dealing with data sets, or people, that don't have say over what goes on.

So, I think about this, for example, with Facebook, right, which is -- and part of to keep in mind of all of this is all of the businesses have different reasons why they're doing different things, right? Facebook wants to give you a service that if you have not signed up to their site before, they want, when you come in, that you don't end up in this weird desert of no friends, no content, no nothing, right, because that's miserable. And so one of the things that they have gotten much better at doing is determining, before you've even shown up, what is the likelihood that you sit within a particular network?

Now, they can do this because of the fact that your friends have most likely updated your email or added your email address to their system, right? So, your friends made decisions to give information about you to Facebook, right? They can do this because they can also assume, once they have that basic information, they can make who else within the network -- what do the people like, what are they interested in, and they can start to say, hey, might you be interested in this, and give you some channel to start engaging.

But -- and this is where we get to this question of -- you know, what kinds of data are we talking about? That individual never gave over their information, they didn't give over their list of friends, their friends gave away them and the site was able to interpolate. And this is what becomes part of the challenge of a lot of the data analytics technics that we're talking about. We're not talking about a known trade-off between an individual and a data analyst. We're talking about the way in which an individual is positioned, intentionally or unintentionally, within this network based on what they have or have not given over, or what's been given over about them without their even realization of it.

MS. ARMSTRONG: So, let's follow this up a little bit. So, how does it -- does it matter how this data's being used? I mean, danah's been talking about the social network context. I'd like to take it back a little bit to traditional marketing or eligibility-type determinations. Does the use of the data help define how it -- how it should be collected or how it should be used?

MR. DUNCAN: Models are at best, as I think it was discussed earlier, just estimates. And we don't know how reliable they're going to be in every instance. And you can imagine -- and they can be accurate or not. You can imagine a company trying to sell a very expensive automobile, and it pulls various lists, and it says there's a 30 percent chance that people will come into your showroom to look at this car versus another list there's a 20 percent chance and five percent. So, they -- they have the money to send out 10,000 solicitations, and they're going to obviously pull from that first list. They might not realize until later that that list is 95 percent men and five percent women.

Now, is that a fair determination? Is that accurate for that car? Well, if the car happens to be, say, a Maserati Gran Turismo, it may turn out that men are much more interested in a car that is a $200,000 phallic symbol than are women.

(Laughter.)

MR. DUNCAN: But you can't really say that the -- the use of the analytics was inappropriate in that case.

MR. ROBINSON: Can I -- I think one thing that is so important and is sort of not yet part of what we're often talking about, but is sort of under the surface of what we're talking about, is the desire that consumers, and historically, the regulatory regimes have to understand why decisions were reached.

So, one of the big things that happens in the Fair Credit Reporting Act (FCRA) context is that if an adverse decision is reached, of course, the consumer has this right to have explained to them why the decision was reached, which means that if new kinds of data are being used to reach FCRA-covered decisions, there needs to be this ability to spell out in some fashion how did that decision arise from that data.

And, relatedly, in the Equal Credit Opportunity Act (ECOA)context, a model that has a factor in it that's correlated with protected status, which, of course, many of the key factors are that predict creditworthiness, sadly, because creditworthiness is itself not uniformly distributed across protected status groups and the majority.

So, how do you decide whether -- notwithstanding the fact that it correlates, say, with race, a factor can still be used in the credit model? And it turns out there's a -- there's a two-factor test. One is that the factor has to have a statistical relationship to creditworthiness, which is unsurprising. And the other -- excuse me -- the other requirement is that the factor has to have an understandable relationship with creditworthiness.

So, under existing ECOA precedent, if buying seeds at the store predicts that you are a bad credit risk, and someone wants to use that in a credit model, even if the prediction is stronger than lots of other more intuitively financial-related factors, it may nonetheless turn out that that use is not acceptable because the relationship is not -- in the words of the financial regulatory guidance -- understandable.

And I actually think that one -- it's a central tension in big data because when you think about the promise of it, it's to surface relationships that weren't intuitively obvious to us in the first place. Things we didn't already know, but then, nonetheless, are useful in the marketplace. But I think that, you know, to the extent that the payoff from these new technologies is to tell us stuff that we couldn't intuitively have figured out, by the same token -- it's a double-edged sword, right -- by that same logic you have the problem of it being very difficult, potentially, to explain either to consumers or to make visible to regulators what the relationships are, or even for the -- the decision makers in business themselves to understand what are the reasons why certain factors are ending up in these models.

MR. GSELL: But there's also a tendency to have big data be more inclusive than exclusive. And I'll give you a quick example. We work with the State of North

Carolina and their education system. And one of the things that has been determined to be very important about education and going through education is the ability to take Algebra in the eighth grade. Okay. Now, historically, the way you got into eighth grade Algebra was teacher recommendations. We've been able to work with North Carolina around analytics to analyze test scores, just pure test scores, from the fourth grade through the eighth grade -- through the seventh grade actually, to determine that there is a group of the population that is normally not considered for pre-Algebra, or for eighth grade Algebra based on combinations of things that are beyond just the test scores, or things in the test that are more than just the actual answers.

And as a result of this, we've identified -- or the State of North Carolina -- the schools have identified 20 percent more students who were not eligible for eighth grade math based on teacher recommendations. And of those 20 percent more students, 97 percent of them go through eighth grade Algebra without a problem. So, they would have otherwise been excluded, but through big data and analytics they're included and they succeed. And it's a huge win for inclusion, not exclusion.

MS. ARMSTRONG: Okay. Let's -- does anyone have some examples of how big data has been inclusive or solved a problem similar to what Gene has laid out in either the traditional credit or marketing/advertising context?

So -- all right.

(Laughter.)

MS. ARMSTRONG: Then let's -- let's take this a slightly different way, but I would like the panelists to be thinking about real examples that they have, because one of the goals of this panel is to sort of lay the landscape of current usage. So, let me --

MR. GSELL: I have lots more.

MS. ARMSTRONG: Oh, good. Well --

MR. GSELL: But I figured I wanted to let other people talk, so I'll just -- I'll hold them and work them in.

MS. ARMSTRONG: How about -- why don't -- you can -- how about why don't you do another one and then we'll see if that triggers.

MR. GSELL: All right. So, along the lines of credit scores and how people are included or excluded, through the use of better data and better analytics, one of the large auto companies that issues credit on a regular basis has been able -- and, historically, they're very conservative, okay, which is we want our risk profile against our consumer loan base to look like this. They've been able to use big data to actually include more people in the sample set than exclude. So, they actually have a mantra, which is how can we be more exclusive, turn down less people if you will, okay, so that we can tease out the people who historically don't have a good FICO score but they are in fact still good credit risks. Okay.

So, working with them and through the analytics we're able to find the people who are normally excluded, include them back into the population to give credit to. And, again, the historic default rate on the incremental people that we bring back into the population is lower than the historic credit failure rate across the entire data set.

MS. ARMSTRONG: So, I think that weaves into one of the comments that David's paper that was released earlier -- or last week -- noted that 70 million consumers do not have credit scores. And that alternative data can often be a positive way to include people that previously aren't part of that mix. So, Gene, without going into the special sauce, can -- can you tell us what kind -- what is it about the scoring and analytics of credit that allows non-traditional data to be used in such a positive way?

MR. GSELL: So, I’m not a credit expert.

MS. ARMSTRONG: Okay.

MR. GSELL: I will preface by telling you that. There’s an ability to get more sophisticated modeling across a larger data set. And the more information I have -- it’s a classic statistical problem -- the more information I have, from a statistical forecasting perspective, the better able I am to predict. So, by bringing in more data, different vehicles, different data vehicles, I’m able to, if you will, tease out, okay, the most likely to be successful credit worthy people. Okay, but I can’t tell you what the algorithm does.

MS. ARMSTRONG: Okay. All right.

MR. ROBINSON: I mean, so, just to go specifically to sort of additional data and credit worthiness, I mean the big sort of frontier there that has been -- that has shown signs of statistical strength has to do with the payment of utility bills, so cell phone bills, power bills, things like that. And, you know, on the one hand there may be people for whom traditional, you know, FICO score data does not exist, nonetheless, they’ve been paying their power bill on time for many years. Turns out that’s a good predictor that they would be a good loan risk. And so, by including that data there is the potential to expand the group of borrowers for whom the lender can have confidence that they are likely to repay.

Nonetheless, when you change how data is used from one purpose to another purpose there are also social justice risks. So, in this context, for example, with utility payments in New England there are many states that have assistance programs where if you are unable to pay your power bill they will keep your heat on in the winter, but what they require you to do is show that you're delinquent in the payment of your power bill in order to receive the needed assistance. They say you don't have to skimp on food, you can buy your groceries and not pay your power bill and then we'll come in and help you. Of course, if the world changes in such a way that that power bill now becomes also the key to accessing credit, then that conflicts with that assistance program in a way that may lead those people to have, you know, a really difficult choice where the state assistance program ends up, in effect, saying that you have to commit some kind of like, you know, credit self-harm in order to keep on getting help keeping the heat on in the winter.

Now, of course, the possibility exists to revise, you know, those programs in ways that resolve that concern. But I guess what I'm really saying is that the -- the benefits that are there, I think, are best realized when we tread particularly carefully with the repurposing of data that was gathered in one context, you know, for use in another. And I would again say the use of data to lock people out of transactions that was at first gathered for market purposes where errors were much less of a concern is a serious social justice concern.

MS. BOYD: So, you'll notice that one of the things that happens is that we're often going to public sector examples. And part of the reason why we do this, even as corporates are working with public sector, is the fact that many of the decisions that are made within private enterprises are not visible. And so, this becomes a trade-off, right. Do you assume that the private sector actors are inherently evil, or do you assume that they're actually trying to do the right thing? And, right, we can agree or disagree on a whole variety of that.

And I think that's actually where it becomes really difficult, because these same technics that can be used to increase different aspects of fairness can also be used to create new kinds of complexities. And it's that tension that becomes really difficult because it's often not visible. And it's not only just not visible to outsiders, it's often not visible to the actors themselves as they're trying to do a lot of the predictive analytics that they're working on. Right. We're working with complex learning algorithms. Do the engineers even understand what's going on? And this is where we get back to this question of scoring as an example there.

Now, the other thing is that when you do this kind of work, what do you do as the intervention? Right. So, I'll give an example. So, in Microsoft Research, which is the academic arm of Microsoft, which is nice because it means researchers publish a lot of their experiments. And so, you can see certain attempts to try to figure these things out. And I'll give an example from a non-focus on discrimination, but it shows the challenge here.

Eric Horvitz is a researcher at Microsoft Research and he's at the point with Bing data where he can predict with a high level of probability, depending on somebody's searches, whether or not they're going to be hospitalized within the next 48 hours. Right. That's a really interesting puzzle. Now, the question is, what do you do with that information? Right.

If you are Microsoft and you are running Bing, does that mean you send a warning sign, like, you're about to be hospitalized. Like, that's creepy, right.

Like, what's going on with that? Does that mean you figure out, you know, a subtler way, a slight advertisement, as a way of suggesting that they might think about it? Again, where do we get on the sort of, you know, Minority Report creepy zone of it all? Or do you not do anything because you don't want to, you know, deal with the liability? Those are ethical questions that become part of it. Things that companies struggle with all the time when they're doing this. Right. They start to see a trend, they start to realize a correlation, and they go, okay, how do we intervene in an appropriate way?

Now, of course, this also becomes a challenge when companies have to think about the responsibility they have beyond their particular domain. So, for example, JP Morgan and Chase does amazing analytics work to predict with high levels of probability whether or not somebody is engaged in trafficking of humans, particularly for sex. Right. And they can do this based on a whole set of different financial patterns that become obvious. Okay. So, their response, because, you know, their company, they don't know how to intervene in human trafficking, right, why should they? So, of course, they're going to work with law enforcement. But that sometimes is a good idea and sometimes not. Right.

And a lot of people who work on trafficking issues have identified why often law enforcement is not the best intervention point where social services is. So, how then do we think about the ethics of those responses?

And this is where we've got this big challenge with corporations. What are they choosing to look at? Are they choosing to do it in a way that we deem to be ethical or appropriate? How do -- what do they do with the information that they get? And when and where do they, or should they make this information public?

And it's not easy to work things out. So, I don't want to assume that just our silence and failure to give examples is not that companies are engaging always in bad -- you know actressing. A lot of is that these things aren't visible for a whole variety of complex ethical concerns.

MS. ARMSTRONG: And I think that's one of the points of Kristin's that the report showed last year. Would you care to elaborate on that?

MS. AMERLING: Yes. We ran into this lack of visibility issue in a number of ways when were looking at the practices of the representative data broker companies. First, the companies are gathering information largely without consumer -- direct interaction with the consumer, so the consumers themselves aren't really aware that the companies are using their information or that the companies necessarily even exist. And then, in looking at the contractual provisions provided to the committee, we saw that that many of the companies perpetuate this secrecy by including contractual provisions in their contracts with their customers that say you're prohibited from disclosing what your data source was.

And then, even when a number of companies do provide -- a number of the companies we surveyed do provide some rights of access for consumers to look at the data that they have on them. And in some cases they provide some rights of correction if the consumer feels the data is inaccurate. But even when those rights are provided, and not all companies do provide them, they don't have much value when the majority of consumers aren't even aware that the companies exist or are collecting this data.

And then, we, in addition, ran into several large companies that outright refused to provide to the committee who were their specific data sources and who are their specific customers. So, those were all obstacles to trying to understand, you know, how the -- how this information is being used and analyzed.

MR. DUNCAN: Companies are in a very

interesting situation right now, especially in the retail community, because we're in a transitional period. For a long time in the world there existed the online community, which a great deal of information tends to be gathered. And then, there's the in-store community where it's a lot more -- a lot more meager. And we've seen a behavioral in stores and in consumers where they want to view this as omnichannel. And they want to buy it online, and they want to return it in the store. Well that means there has to be data flows back and forth between those two -- those two markets. And so, the folks who are running the store have to figure out how far can we go?

And what we find happens -- and this may explain some of the information shortages that you're talking about -- what happens is that they look at correlates to what consumers expect in terms of the use of information in the store, and that's the model they use. So, they tend to be very conservative in terms of expanding the use of the data or the expansion of that data in a store market.

MS. ARMSTRONG: Can you give an example of that?

MR. DUNCAN: There is -- there is what -- there may be cookies that are used online that will travel from location to location. In a store environment we're uncomfortable with that kind of movement. We would say consumers are comfortable being observed in the store, and so information may be gathered and used within the store context. But they're very reluctant to go beyond that because that violates consumer's reasonable -- or the -- that violates the store's expectation as the consumer's reasonable expectation.

MS. BOYD: Let's be clear that Mallory's hinting at the fact that there are actually a lot of startups out there that are actually trying to track mobile phones into stores. And there's a big tension within the retailers as to whether or not to implement that because it parallels the cookies issue. It allows you to literally track a unique identifier of a phone, see whether you've seen that person before, see what their patterns are, see how they're navigating the store, all of that is technically feasible, the question is whether or not retailers want to implement it or what the challenges are of doing so.

MS. ARMSTRONG: I think Joe wants to add something.

MR. TUROW: Well, I've spoken to a couple people who say they do exactly that now. And all you have to do is think about loyalty cards. Loyalty cards, which are kept by virtually everyone here who goes to a supermarket, probably uses a loyalty card, it's like 90 percent of Americans who go to supermarkets that give out loyalty cards use them, because otherwise you lose a lot of money if you don't. They track everything you do. Until the last few years they haven't been able to much with it, they haven't, for lots of reasons, done any big data analysis, and that's changing totally. Okay. And there are companies, for example, Kroger owns part of Dunnhumby, which is a company that is designed just to do these sort of analytics. The idea now -- companies like Macy's and others are putting pods of these beacons in stores that look at you when you reach you a certain point and then give you specific blandishments, like, discounts based upon your shopping habits. Catalina Marketing for decades have been giving people these long coupons as you check out, based upon 52 weeks of looking at your shopping habits anonymously. Now they're beginning to do stuff in the store in a digital sense and outside the store.

So, we -- in fact, you're absolutely right what's happening now is stores are getting so nervous about the online environment that physical stores are bringing the internet to the store. And the big data are extremely a part of that in ways that danah mentioned and in other ways as well. And it’s a -- that’s exactly what’s happening. It’s a fascinating trajectory partly because of the growth of big data in the online world.

MR. DUNCAN: And, if I could, it’s also because the consumer expects that seamless experience. And it presents the retailer with a bit of a dilemma. You want to treat the consumers in the way they want -- like to be treated, but you want to be sensitive to the privacy implications and the use of the data at the same time. And how you square that circle depends on the reputation of each retailer.

MS. ARMSTRONG: But is it a transparency issue? I mean, do you think we’re at a -- that in five, ten years it will be totally different because the consumer’s expectation of privacy or not sort of being their purchases or their behavior being followed? I mean, I almost hear you saying that it’s sort of expected online but not in a store. That seems like a little bit of a disconnect to me.

MR. GSELL: Well, to some extent, it’s generational.

MS. ARMSTRONG: Uh-huh.

MR. GSELL: So, I mean, I am high on the creep factor --

MS. ARMSTRONG: I was going to say, you and me are the same generation.

MR. GSELL: -- on some of those particular things. Yeah, but my kids, you know, they have no problem.

MS. ARMSTRONG: Right.

MR. GSELL: They expect that to your point. They expect the same kind of offers and service and interaction online when they walk through the store they expect the same experience.

MS. BOYD: Now, I think I'd be -- I want to sort of butt in there, because young people -- there's a lot of self delusion. Young people are actually just as self deluded about a lot of this as we adults are. Like, there's not this big difference between young people. They want privacy, too. They're focused very heavily on the people who hold immediate power over them.

I want to just think through an experience all of us had. Right. We came in here this morning, in some ways we knew it was going to be recorded, we knew people we're going to take pictures, we're at a public event, right. You saw the webcast notice. And yet, when we heard this morning the listed detail of, like, if, you know, if you object at any moment to a photograph being taken, you know, as Tiffany went through this you're sitting here going, "I want to leave," right, like, "This is really creepy." Right. And even though you know it part of it is that you had put it down, you had avoided it, you hadn't thought about your hair in perfect, you know, coiffed form.

This is one of the challenges that we run into all of the time, which is that notice and information is not always the best way to actually create a meaningful relationship. And there's a lot of self delusion on both sides. The reality we also -- we collect a lot of videotape that we never look at. Right. My guess is that most of us are never going to look at the videotape of how badly our hair looks on that camera. Right. Part of it is this interesting challenge of how much do we purposefully sort of put this information aside and navigate it through.

But I would not put this as a generational issue. This is not a generational issue. And Chris Hoofnogel, in particular, has done phenomenal work looking at the consumer side of it. Young people feel the same way as adults, their trade-offs look different.

MS. ARMSTRONG: But is it an educational issue, then? I mean, it's easy to suggest that it could be a generational thing or not, but I -- I wonder how do we educate people, not just adults, not just children or younger people, to expect that or to know that their transactions will be recorded or collected.

MS. BOYD: But you're basically asking to educate them about the fact that they are powerless. Right. Like, that's what the education ends up being about. Like, either you opt out of this room, right, or you'll be recorded. Period. You have no say. And that's one of the trade-offs that happens all the time online, or in these -- you know, commercial environments. Right. You want to go and buy something from Best Buy, you will be recorded, get over it. Right. Otherwise, don't go into Best Buy.

MR. ROBINSON: And just to pick up on this transparency and on something that danah earlier said about how, you know, we go to these public sectors examples because we don't know what's going on inside of these private enterprises. I think that's absolutely true and is central, really, to the FTC's future decisions about what to do in this area, is that, you know, what -- education about the fact that a practice happens in general does really little, if any, help to try and figure out whether that practice manifests in a discriminatory fashion for particular people.

And Dr. Sweeney's work on the discriminatory delivery of online ads is indeed an unique example available in the public discussion, which is why the

Chairwoman mentioned it this morning and we've come back to it here. And I think what I'd like to see is a world in which you don't have to be a -- you know, a world leading data scientist, who also happens to personally be the victim of discrimination, in order to have the tools that are necessary to check that that's happening and address it. And certainly after the study came out, Google changed its practices with respect to the delivery ads opposite names in general in order to avoid the discrimination harm of these disparaging arrest-suggestive ads.

But that's an extremely unusual case and I -- and I think we would all like to see a world in which if harms like that are happening to people who, you know, are not academics and data scientists with kind of all of the resources that it would take to be a personal, you know, sort of scholar of that discriminatory harm, you know, when that harm befalls someone who's in a different position, who's more in a marginalized position, I think what we would all like to see is for those harms to be treated with equal seriousness. But I think the fear that the community has right now, which I think is an extremely well-grounded one, is that when harms of that sort do befall someone who's in a marginalized position, they really don't have the tools today to -- not only to solve, even necessarily, to diagnose those problems.

MR. DUNCAN: It’s not --

MS. ARMSTRONG: But some -- sorry, I was going to say that some would argue that the Fair Credit Reporting Act is a -- is a mechanism in the credit context, because it’s doing exactly the sorts of things you’re talking about which is when adverse action -- if you fall within, an adverse action is taken, you’re provided a notice that the adverse action was a result of something in the credit report, and you’re given the opportunity to dispute that information. So, I wonder whether the expectation in the credit world is a little bit different because they know they have this mechanism in place, and whether that’s a metric that’s useful in another context?

MR. DUNCAN: I think we have to make qualitative differences. When we’re talking about credit, or insurance, or education, we may have very different expectations than when we’re talking about marketing.

Let me go back a moment ago to the example of the sports car. One solution would be to say, no, you must send the offer to come in and test drive the car to more people. Well, the consequences to that is that people receive the offer who have no interest in, thus depleting the funds that the dealership has for sending it out, or people will rush in to test drive it who have no ability to purchase the car, thus tying up the service folks at the auto dealership.

So, you really have to look at the quality of what you're doing as opposed to just saying let's take the credit reporting structure and apply that more broadly.

MS. BOYD: Also, I don't want to dismiss the credit reporting. I think it's an important intervention, and I think -- you know, I'm very excited to see that being a regulatory intervention. But also, let's be realistic. Many of the people that are most hit by it have not the time, not the connections, not the understanding, not the literacy, not the wherewithal, and they don't feel a sense of power to be able to actually fight it in many cases.

And so, when we actually look at that, it's also this question of who has all of those resources, those soft resources, to be able to do the thing that they were supposedly protected, you know, for. And that's where this interesting tension emerges of where are we trying to get marginalized voices, whether we're talking about youth, whether we're talking about protected classes, to raise up and try to be powerful against systems of power that are meant to actually challenge them? Or where are we trying to think about the role of different kinds of advocacy groups or different kind of actors who work on their behalf? And I think we have to be realistic about how we're dealing with this.

This is the challenge with education. I think a lot of our education narratives go back to consumers without actually thinking about the lack of other resources that they have to make sense of, or feel agency or power in light of what's going on. And I think that's a difference between how we think about it theoretically and what we think about in a regulatory context, versus what I see on the ground, when I deal with a lot of marginalized people who are just like, I don't feel like I have any sense of power to do anything about this so don't tell me about it.

MS. ARMSTRONG: So, what's the solution? What are your recommendations for empowering those people?

MS. BOYD: I mean, this is where I do believe -- I believe strongly in the role of advocacy as a mechanism to be speaking on behalf of groups. And this is one of the reasons, you know, Dave and I spend a lot of time talking with different legacy civil rights groups for this reason. Like, those folks need to be educated, you know, on behalf of populations as opposed to -- and they need to have the transparency and the tools and the mechanisms with which to hold, you know, systems of power accountable without always going direct to the consumer as the right direction there.

MR. ROBINSON: I mean, so, I mean, these are groups that have unique -- you know, that hold the franchise through their -- and have earned the franchise to speak for these communities and policy settings. Right. There are people who -- whose job that is. There are people who do it for, you know, down to migrant farm workers, and really the most marginalized, you know, people, you know, in our country have, you know, people who are there.

But I think making the practices transparent enough to give handholds to advocates in those cases in which there’s a role that they do need to play I think is a role that FTC itself has often successfully played. And certainly, I think the FCRA is, you know, a good model for the things that it applies to and has certainly -- has played a role in making underwriting a relatively conservative area in terms of the applications of big data as compared to these unregulated, you know, marketing practices.

Although, as the Chairwoman noted in the case of these thinly aggregated scores that may be used to lower credit limit that are putatively outside of FCRA, I think it becomes difficult. And frankly, I think there are, you know, legislative and ultimately constitutional questions about how far the FCRA-style model could be extended into the marketing world that I think really do force us to -- and I also -- let's -- you know, law and regulation have a valuable role to play, but so does -- but so does corporate citizenship potentially. I mean, I think people who say, you know, we're doing stuff in a way that we would like to be responsible and we would like to take affirmative steps to make sure that we're not inadvertently having disproportionate adverse, you know, impacts, I think there's a role actually there for collaboration with advocates. Because right now it's clear what the sign posts are, what the benchmarks are for making sure that you're not doing these things inadvertently. And I think that if I were to project forward five or ten years, my recommendation, my hope, and also my prediction, would be that there are going to be some practices that emerge, and my guess is that they are going to emerge probably in a collaborative fashion that's probably outside of the legislative process.

MR. DUNCAN: David, I want to be very careful I think here, because access to credit is essentially a fundamental right in this country. Access to a high-end men's fashion catalog is not. And we ought not to conflate the two in this discussion.

MS. AMERLING: But the --

MS. ARMSTRONG: Well -- go ahead, Kristin.

MS. AMERLING: I mean, the -- the kinds of products that we saw in our review of data broker practices that involve marketing did go beyond products designed to promote the most appropriate car or reach the people who are most interested in cooking magazines. I mean, there are a wide variety of groupings of consumers based on their financial and house status that includes lists of people who have diabetes, Alzheimer's, or suffering from depression that consumers may not be as happy to find that they're on as finding out that they can be targeted for the best car that's most tailored to their needs.

And there was actually an interesting article that just came out last week by Bloomberg on widespread sale of health ailments list that goes right to this point where they reported that just with simple Google searches the reporters were able to find lists of consumers with their names and addresses that were identified as associated with specific diseases. And they interviewed some of these consumers, and one who was associated with diabetes was surprised and not at all happy to find out that he was on this list, and said he didn't have diabetes and nobody in his family had it.

So, there are some sensitivities raised by some of these products that I think are a little more in the grey area than just these are the best products to tailor to the needs.

MS. ARMSTRONG: So, we're about to run out of time, but I'd like to give everybody on the panel an opportunity to say some parting remarks. We have some question cards from the audience that raise some issues that I think would be worth mentioning. And that is the level of trust that may appear to be missing in the big data context of the relationship of marketers, a person that goes to a store may choose to go to the store, there may be a level of trust there, but the invisibility of big data disperses that trust a little bit perhaps.

But I would -- I would like each of you -- and I feel terrible in a way because we have ended this panel talking about what the last panel is going to be talking about more, which is sort of the path forward. So, as you provide your final little remarks, if you would also remember that we were laying the landscape and if you could bring it back to what's happening now as we wrap up, that would be fabulous.

MR. TUROW: Okay, I -- I had a path forward, but I'll try to make it a now.

MS. ARMSTRONG: As long as you bring it back to the landscape.

MR. TUROW: The now part of it reminds me about the -- I think it's shameful that in a commerce committee hearing when a senator asks a representative of the data industry whether he could name his clients, he refuses to do that. These are areas of life that impact all of us. And the collection of information about us and their use, I think should be required -- I think companies should be required to say which data broker -- the data broker should be required to say what -- who they get it from, what are the categories, because these affect us everyday.

In terms of education, I think most people are learning about credit cards and loyalty from Jennifer Garner on tv commercials then they learn from anywhere else. We have no learning about this stuff anywhere. People are -- it's totally obscure. And I would suggest that's purposeful.

I think the idea of big data is a continuity. There's an element of continuity between that and the quantification of the individual that has gone back 30, 40 years. But we're in a century now that I think will be looked at as the century of data, the century of pinning numbers on people and trying to figure out where that leads people. And we're only at the beginning. So, I think we have to realize that this stuff is important, not just for now, and it's going to get much stronger with greater processing and the kinds of things that people are saying today, "we can't do it," are going to be done.

So, the issue is not, you know, is this going to happen because it's too futuristic, but when it happens are we going to have the conceptual tools to deal with it.

MR. ROBINSON: I just -- to sort of pick up on the question about trust and where things are today, I think there's an unrealized opportunity to create greater trust with consumers in terms of how these technologies are being used. And I think that the tools that we have from prior regimes about notice that your data is being collected -- the notice and content regime, frankly, I don't think offer the tools to create that greater trust. Because, as danah was saying, the data is collected in a way that you don't have fine-grained awareness, and you certainly don't have fine-grained choice about what's going to happen.

And I think that the tools that we need in order to be able to have practices happen that -- that gain the predictive payoff from these analytics, but at the same time give consumers good reason to trust that things are being done in a way that they can feel comfortable about, I think those tools have really not been perfected yet, and that we're in a place -- we're in an exploratory initial place now of needing to build new tools for accountability and trust consistent with the business leveraging of these -- of these tools.

MR. GSELL: I guess what I'd say is the genie's out of the bottle. Stuffing it back in is not going to happen. Data is a part of what's going on. There's more of it than there every was, and there will continue to be more than there was last year or this year.

I think that, for the most part, the uses of it are much more positive than negative. There are enormous examples of big data being applied to solve big problems, big worldly problems, big human problems, and healthcare, and in genetics and in disease control, in commerce in terms of how to minimize fuel consumption across airlines or UPS, or people like that. For the most part, it's really very, very positive that we can now compute on data that wasn't even available two, three, five, ten years ago.

From a consumer perspective, again, I think the economic model still will drive most of the thought process around this. A retailer doesn't want to do something that creeps you out. Okay. And the minute they cross the line they get what is the worst thing possible for them, which is you opt out. And the worst thing for a retailer is a fair amount of opt outs. They want to keep you in the fold. They want to be relevant to you. They want you to be responsive. And their only notion is to give you something more relevant to you so you don't have to filter out all of the noise that's out there.

I think that there are clearly some privacy things that need to be monitored and watched, but on balance I think most consumers are electing to opt in as opposed to opt out.

MR. DUNCAN: I think Gene said it well. I mean, there are a lot of retailers out there, several million. And so, there's a lot of choice and opportunity for consumers. And trust, in that context, is more than just one element, such as sharing this data flow or another, it really is about developing loyalty with the customer so the customer trusts the retailer and wants to return and maintain that loyalty.

One easy example. There are companies out there that gather -- like, Amazon -- gather huge amounts of data, and yet, consumers know this because they see the sign that says if you like this item, you may like that item. They appreciate that, and they go back and shop again and again, because they trust Amazon to do what's right by them. And that's what other stores are aiming for.

MS. BOYD: My perspective of this space is actually extraordinary complex, and it's not that they're not inherently good actors and evil actors, it's the fact that everything has a lot of grey zone. You know, the other thing I think is important to highlight in this is that we often talk about companies that we're thinking about as high-level brands. Brands that we can hold accountable and recognize. But then we also deal with data brokers whose names nobody recognizes who are holding on to data, who are buying data at bankruptcy situations, who are capturing things that -- you know, and pulling together data sources that we don't even know about. And this is one of the reasons why this space gets very murky because we often talk about it within specific silos rather than the complexity of it.

Anne Washington's been talking a lot about data supply chains, which I think is a way of interestingly thinking about it. It's a metaphor. It's not a perfect metaphor, but it's a really interesting metaphor to start thinking about that. How do we start thinking about holding supply chains accountable when we're thinking about these data issues? Not just in terms of the data brokers that the FTC is looking at, but in terms of all our own acts -- our own behaviors around this.

The other thing I think is really important to highlight is that many of the companies, especially the big names, are really trying to do their best. Right. They're trying to figure out how to hold this stuff in a responsible way. But as, you know, David's point out, they don't always know what the best practices should be. And this is where there's tremendous opportunity for meaningful cross-sector collaboration to try to figure these things out.

Regulation is one approach. It's a very power strong-armed approached, but collaboration is another approach to start thinking about how do we evolve the best practices and how do they differ per sector, because as Mallory pointed out it's different when we're talking about retailers than versus what we're talking about in terms of finance and credit. What does it look like and how do we pull things together?

Finally, I want to sort of end with a philosophical point, which I think is also about the state of being. The notion of a fact in a legal sense emerged in the 1890s. It's a really modern concept. And anybody who lived through the last election in this country saw that we're kind of in post-fact state.

(Laughter.)

MS. BOYD: For better or worse, one of the things that's sort of coming up as a new equivalent of fact is rethinking probabilistic understandings. This is the big data element. This stuff is here to stay. Part of it is understanding what probabilistic systems mean for our whole ecosystem, because part -- in understanding probabilistic systems you realize it's not cleanly fact, it's about trying to figure out how to deal with this, and how do you hold probabilistic systems accountable, and how do you think about their role in things like rule of law is going to be very, very messy. And this is where I say this because a lot of what we're dealing with in terms of the systems that we're trying to hold accountable are probabilistic systems, which are not intended or designed to be discriminatory in a traditional sense in the narrative of a fact, but they're done in this way that ends up unintentionally doing so. And that goes back to Solon's comment. And I think it's really important to understand that philosophically, because that's one of the things that we need broad-spread literacy on before we run into the systems where we just assume to treat these things as facts.

MS. AMERLING: I just want to go back to the issue of transparency and visibility. That's a theme that emerged from inquiry; it's emerged in many of the comments today. The Chairman has proposed legislation to provide consumers access the right to correct their records, the right to opt out if they don't want their information being used for marketing, and this is kind of a baseline for transparency and it's very interesting to hear about these additional non-legislative tools. We recognize this is a complex and evolving issue and are looking forward to continuing to being part of the dialogue about the impact of big data on consumers.

MS. ARMSTRONG: Well, I wanted to -- I want to thank everybody for participating in this panel and bringing the different perspectives that you have. I think one thing that seems fairly clear is that there is no single solution or there's not even any single way to look at this. That it's very much something that we must look at through a multi-faceted lens when we're talking about marketing, credit, social media, and all these other topics.

I hope we we're a little successful in laying -- assessing the current environment, but I know that the panelists here could have actually participated on any of the panels today because it all does, as danah said, a lot of grey areas. So, thank you very much everyone.

(Applause.)

MS. ARMSTRONG: And you need to return -- audience members, you need to return here at 11:00. You have about a ten minute break. There is a cafeteria, but you can’t bring any food in here, so...

(Laughter.)

(Whereupon, a brief recess was taken.)

PANEL 2: WHAT’S ON THE HORIZON WITH BIG DATA?

MS. GEORGE: Hello, welcome back. We’re going to get started in a couple of minutes. Will the panelists on the second panel please come up to the stand? Please take your seats.

(Brief pause.)

MS. GEORGE: Good morning again. For those of you who may have missed the beginning, my name is Tiffany George, and I am an attorney in the Division of Privacy and Identity Protection here in the FTC. And welcome to our second panel. We’re going to discuss what’s on the horizon with big data. As you can see, the first panel touched on a lot of different issues, some of which will be covered in our subsequent panels. But for this panel, we want to focus on potential future trends in big data practices and implications for consumers and organizations.

I'd like to thank our esteemed panelists for joining us today. I will briefly introduce them and then we'll dive right into the discussion.

Joining us today are Alessandro Acquisti, Associate Professor of Information Systems and Public Policy at the Heinz College of Carnegie Mellon University and Co-director of the CMU Center for Behavioral Decision Research; Pamela Dixon, founder and Executive Director of the World Privacy Forum; Cynthia Dwork, distinguished scientist from Microsoft Research; Mark MacCarthy, Vice President for Public Policy of the Software Information Industry Association; Stuart Pratt, President and CEO of the Consumer Data Industry Association; and Nicol Turner-Lee, Vice President and Chief Research and Policy Officer for the Minority Media and Telecommunications Council.

Welcome and thank you again for joining us.

I'll start with a broad topic for our discussion today and then we can drill down. So, I'll toss this out to the entire panel. What trends do you see in the future of big data? Is it going to get bigger? Is it going to be better? Will there be more passive collection of data versus active collection of data? How will it be used, such as for marketing, fraud detection or eligibility determinations? And should consumers be concerned about these practices?

MR. MACCARTHY: Let me jump in. Is the mic on? Can you all hear me?

AUDIENCE: Yes.

MR. MACCARTHY: Good. So, I first want to do some marketing. Our friends at the Future of Privacy Forum and the Anti-Defamation League have published a nice little collection of examples where big data is used for empowering people and promoting economic and social opportunity. I urge you all to take a look at it and contemplate the advantages, the benefits of using big data in many of these contexts.

The couple of examples I want to mention, one of them has already been mentioned, alternative data scores, I think these are going to increase going into the future. A recent study by LexisNexis found that 41 percent of Hispanics and African Americans could not be scored by traditional systems, while only 24 percent of the general population could not be scored. That's an unscorable rate for minority populations almost twice the general population.

Their new risk view scoring methodology allows 81 percent of the people who are not scored to receive a score and thereby be eligible for the mainstream financial products. That's one example. You heard a little bit about that before, but I wanted to put that one on the table as well.

Cognitive computing in healthcare, IBM has a version of its Watson computer that functions as an oncology diagnosis and treatment advisor. It's in use today at Memorial Sloan Kettering and MD Anderson Mayo Clinic is using it to select subjects for clinical trials.

So, how does this help the under-served? Well, there are shortages of specialty providers in hospitals all over the country. Some 50 to 60 percent of community hospitals do not have an oncologist on staff. But now suppose that the medical insights from these computing systems can be made available to clinicians in community hospitals throughout the country. This isn't happening today; it's a potential for the future and it's one I think we should encourage.

The last example was one that was also mentioned on the last panel. These are predictive analytics in education. Many schools are using predictive analytics tools to find students who are at risk of dropping out so that they can engage in early intervention operations. Many companies provide these kind of tools. They're very, very effective. If they're deployed in time, they can reduce the dropout rate significantly.

So, three examples of the use of big data analytics for productive and for socially beneficial purposes that have the effect of increasing social and economic opportunity. We'll have a further discussion about all of these, I'm sure, as we go on.

MR. ACQUISTI: Okay, I'll do some marketing as well like Mark just did. Curtis Taylor is an economist at Duke, and Liad Wagman, an economist at Northwestern, and I just finished a manuscript reviewing the economics of personal data and the economics of privacy.

So, it was interesting, this exercise we did, because we were looking to see what economists over the last 20 or so years have said about the impact that personal information and the trade of personal information can have on the welfare and allocation of surplus. Because, to me, going back to your question about what is the next big issue -- for me, as an economist, the next big issue is to what extent the data will increase the economic pie, will lead to more economic growth, benefitting everyone. So, a win-win. And to what extent instead will simply affect the allocation of surplus. So, winners and losers.

The economic pie remains the same. But some entities gain more of the pie and some entities gain few.

So, for an economist, that’s a problem of welfare and allocation. And what we found in the detailed chart is that, well, generally, with more information, economic growth goes up, you have more efficiency and that is predictable, I would say. But there are also cases where paradoxically or surprisingly it’s actually privacy which can lead to more economic growth.

One case in point is health privacy legislation, which can paradoxically promote innovation in the field of HIE, health information exchanges, promoting the growth of HIEs, because it decreases privacy concerns and uncertainty that firms or health organizations may have in terms of how to use their data.

In terms of the allocative effect, we find evidence of, of course, both privacy and lack of privacy affecting winners and losers. Sometimes it’s the transfer of wealth from data subjects to data holders, for instance, the case of price discrimination. Sometimes it’s an issue of transfer of wealth between different data subjects.

One experiment that we actually ran at CMU -- maybe I’ll mention more about it later -- was about the role that personal information found on social media can have on the hiring behavior of firms. And what we find is that even when candidates have identical educational and professional backgrounds there is an impact on the personal information, protected traits such as religion affiliation or sexual orientation in how employers make decisions.

So, this personal data, which employers can find online, can paradoxically create less fairness. So, we have more data, but less fairness. We have, of course, also cases of more data, more fairness, which I believe Cynthia will discuss.

So, the point being that going back and echoing some of the remarks Chairwoman Ramirez said this morning, not only I believe thst, as she pointed out, big data will probably have both positive and negative consequences, but I also believe that market forces alone will not necessarily weed out the bad from the good, because what we see in the literature is that market forces can create both the bad and the good.

MS. DWORK: Can I jump in here? This is not advertising. Maybe it’s a call to arms. So, instead of answering the question of what trend do I see, here’s a trend I would like to see. I would like to see big data being used to detect discrimination. I’d like to see big data being used to find ways of countering discrimination. I’d like to see big data being used to analyze how people behave and know how to make suggestions to make their lives better. And much of the talk on the previous panel was somewhat defeatist in this regard.

And I think that danah is right, that we need advocacy. We need somebody who has an interest in it. If we rely only on people who have a financial well-being, how are they going to get organized, in this particular case, to help themselves?

MS. DIXON: Hi, thank you to the FTC for the invitation. I appreciate the opportunity to talk about this issue, which is very near and dear to my heart.

So, I've really thought about this issue an awful lot, for a lot of years now, and early this year Bob Gellman and I put out a report called The Scoring of America, and a lot of our thoughts are distilled into those 90 pages. And it took 90 pages because big data is really in a formative phase right now and there are a lot of signposts that point to this. But I want to really dig at the root of the matter here and start there and in my comments today move forward from that.

But, to me, the root of the matter is this -- and we really see a lot of things hedging around this, but never really diving down and getting to it. So, to get to it is this: The moment that a person, an individual, is put into a category or is classified in some way or is scored in some way, that triggers a data paradox. We can talk about it all we want and I'm happy to talk about it with you for hours. I can tell you many examples where "big data" has been used to help consumers. I can also give you examples where the exact same data has been used to hurt consumers. And that is the data paradox. If you're a scientist, you may call it the classification effect.

But bottom line, when you classify an individual, you trigger this. And when that is triggered, we have to do something about that in terms of fairness structures. And one of the very big question is what do we do.

So, if you look, for example, at victims of domestic violence, so in order to assist victims of crime and domestic violence, they are put into a classification as a victim of that crime. But if you talk to individuals who are victims of these crimes, they don't want to be in that classification because that reaps some very difficult probabilistic analysis down the road and they feel the effects of that, for example, when they pay higher health insurance rates because they've been the victim of a crime and they're assigned statistical risk.

People who have diseases and rare diseases and chronic health problems have the same problem. So, at the same time, you can use the information to suppress, to lead, to help, to heal, to hurt. So, how we solve that problem of that data paradox is going to be really what we need to get at moving forward in big data.

DR. TURNER-LEE: Thank you to the FTC for having me here at this conversation and to all of you for attending.

So, I want to jump in. I think a lot of people have already said some of the things that I want to say, but I want to answer Tiffany's questions around trends in the future of big data. Is it going to get bigger and is it going to get better? And I want to say, yes, yes and yes. I mean, every day we get out -- you know, I'm sure it was said on the first panel, but every day we get tons of data, individual bits of data collected about us that goes into a dossier or portfolio that, in some way, has an impact. And for social scientists like myself, who my own plug is just working on a paper on privacy and minorities, we don't know where that data is going in terms of its social benefit, but, nonetheless, it's being collected and it's being collected in an exponential manner.

I just attended a brief conference on the internet of things and Cisco has basically stated that the U.S. has a $4.6 trillion stake in the internet of things and the internet of things will only be successful the more data that we collect around the use of those devices.

It’s interesting when I think about data analytics -- and I recently participated in a panel where the question was, is there a good purpose for big data and data analytics and data science? Clearly -- and at MMTC, we represent under-served communities, particularly minorities and other vulnerable populations -- data analytics can certainly generate a social and community benefit. When I think about healthcare and how it can contribute to that -- I know we’ll talk a little bit about that, so I won’t go too far into it or educational outputs of value -- big data can, in some way, actually help us solve social problems related to health disparities, educational disparities, consumer -- disproportionate consumer impacts, et cetera, environmental causes.

One of the examples that I commonly use is when you look at smart meters and low-income communities where people tend to pay higher in terms of their rates, there’s a potential for big data to help us understand better how to preserve income in the pockets of people who are, you know, economically depressed. But at the same time, create healthier communities and more sustainable communities.

All that is great, right? Even with education, there’s the opportunity to adapt the technologies and I think some of the things you talked about in terms of predictive analytics, to help us to better educate low-income minority kids. Again, that’s all great.

But as I said on a panel earlier or last week Mark was on the panel with me -- the data must be protected and aggregated in such a way because, oftentimes minority groups are holding on so tight to the one asset that they have, which is their identity, and we often see that if improperly used -- and I think Alessandro’s paper was actually very good -- we can see bouts of discriminatory behavior that actually impacts it negatively.

So, take the energy example that I just gave, whereas big data could be used for the purpose of building more sustainable communities, it can also be used to tell low-income people that you’re not using your energy too smart and possibly there’s an opportunity for a surcharge. Whereas predictive analytics in education can actually be a good thing to help educators teach better and parents be more engaged, it also suffers the possibility of redlining students in the classroom.

So, we have to think really carefully about this. And we, at MMTC, constantly struggle because we see the value of innovation and what it's actually done in this society, while at the same time, for disproportionately minority, senior, low-income vulnerable populations, the question is can big data produce a social benefit without having a subsequent harm on those communities that are contributing to this. And we've seen, particularly the FTC, examples where some of those -- and I'm sure we're going to talk about it more on the panel because we talked that we would -- but we've seen examples where that discriminatory behavior has a short-term impact and what we fear is a longer term impact when it comes to civil rights.

MS. GEORGE: Stuart, I'm sure you have something you want to say.

MR. PRATT: Yes. So, I was invited late to this panel, so I missed the conference call. And Maneesha called me and said, Stuart, we'd like to have you on a panel, but we've already held the conference call. And, so, I guess I get to say whatever I want because I'm not bounded by whatever was on the conference call. No, so, but I was on an alternative scoring panel earlier this year -- Pam and I were on the panel together -- and I'm glad to be back again.

Joe, I'm missing you here on the panel. So you were on the first one, and taking good notes.

So, I love this dialogue. It's a really, really important dialogue. It's really important that we wrestle with fairness and fair treatment. And that's true for industry organizations, that's true for academics, that's true for some of the nation's largest and most successful companies in the United States. And you've got a great sort of cross section of interests at a table like this. And, candidly, really the best hope we have coming out of this is that we don't just sit on this panel facing outwards, but some day we're sort of sitting around the table looking at each other and having more of that dialogue.

But, Tiffany, thanks for pulling this panel together and for leading our discussion.

So, CDIA is much more -- our members, as the Consumer Data Industry Association, we're much more focused on risk management. So, it's a -- we often are operating data systems, databases, which are a little closer to laws we have on the books today and we're a little further away, if you will, from the question of how you categorize consumers in order to reach them with the right offer. There's some of that. But we're more often dealing with and pushing data into the transaction with regard to how am I treated once I'm heading into that transaction.

So, for example, the Equal Credit Opportunity Act, very important law which addresses core fairness questions relative to credit, of course. The Fair Housing Act, which addresses core questions relative to how I’m treated. But by the way, interestingly enough, both ECOA and Fair Housing also address, to some extent, advertising. They have implications for what do I say when I advertise, where do I advertise. So, there are implications. Certainly, current laws wrap around at least some of the dialogue that we listened to -- and I thought it was a great, you know, first panel -- but those laws are out there today.

And I do think that that’s part of the analysis going forward. You know, how do current laws address fairness and how sufficiently protective are they in some of these transactions? Because our members are involved in a telecom company’s approval of a consumer, an insurance company’s underwriting a decision, a lender’s decision to make a -- what we’ll call a risk-based offer of credit and, of course, we’ve talked a lot about credit scores and they’re a rank ordering system. And, in fact, we think it’s a very effective rank ordering system and it’s important for us to have systems that rank order risk.

Why is that? In the United States, we might lean towards safety and soundness because, in fact, the great recession would tell us safety and soundness is a whole lot more important than maybe we ever thought and we actually could break the system here in the United States and we got pretty close to it.

If you go to Europe, they would say credit reporting systems, data systems like those that the CDIA speaks for, are very important because we want to make sure consumers have the ability to pay, that there's a responsibility associated with the loan or the offer that you make to make sure that it isn't just going to work for you, but it's going to work for both of you in the contract, that the consumer is also successful and it's a good match. So, I think data is best when it's matching the consumer with a -- not just any offer, not just an offer I'm interested in, but an offer that I'm going to be successful in accepting and working with going forward.

That's a little idealistic. I'm not sure we're 100 percent there. I see Nicol leaning in towards me here like this. But I don't know that we're 100 percent there, but that's kind of the promise that we have. But, for us, it might also be a great example real world would be, we think more often now about not -- certainly, many protected classes of consumers through the Civil Rights Act and, by definition, through ECOA and other similar laws, insurance commissioners at the state level, but it's also about identifying consumers whose behaviors have changed because of the economy.

My grandparents lived through the -- you know, the failure really, not just a recession, but a full-blown depression, and you could see the behaviors that they had. But you know what, I look at my sons going through college now and young people that we hire in our offices and we look at the fact that debit card transactions have overtaken credit card transactions, we see some shifts in demographic behavior in databases today.

And so, what -- let's just take a credit report. What a credit report looked like at one time may look different going forward and how we inform the dialogue of a risk-based decision may look different going forward. The fact that I own something may be more important going forward than how I pay a credit card transaction or we need to bring new data in to thicken up systems and to create more inclusion.

So, I do like the idea, though, that there are market forces, which are lining up pretty nicely, with a societal interest, deep societal values that we have in this country. And that is we want fairness, we want equity, we want equality, we want the right treatment for the right consumer. And there's an interest in doing that because it’s, you know, sometimes, to some extent, law, but also because of market interest. This broadens our markets for consumers to engage in a successful product.

And, again, it’s the 50 or 60 million sometimes called credit invisibles in this country. How do we reach them? Well, we need public record data sources. We need utility information because some consumers pay utilities, but they may not be paying on a credit account of some type. We need telecom because telecom is ubiquitous and deeply penetrated into communities of color in this country and used properly, used wisely, used effectively, used fairly. These systems are the kind of systems these data sets and the analytical tools to back them up are going to empower consumers and we will push deeper, but successfully into these markets, successful for those communities and also successful for sort of economic benefits very broadly. So, food for thought.

MS. DIXON: So, to pick up on Stuart’s comments, the -- actually having you on this panel, I think it’s a great idea because regulated industries are already using little bits and pieces of things that are working, such as the Fair Credit Reporting Act and, for example, HIPAA and folks who are regulated by the common rule, people who are doing human subject research.

So, there are pieces that are working, and we've learned a tremendous amount about certain statistical populations because of the credit report and credit scores and the 50 years of history that we have there. Now that it's more public, we know more and consumers can also benefit from that knowledge.

But I want to pick up on something that Stuart was talking about, which is factors. So, let's say that we have the Equal Credit Opportunity Act, and it has narrow applicability, but factors such as race, whether or not you're married, things that really matter in those, you know, financial decisioning processes, they matter in other decisions, too. And when you look at large rich data sets, it's really a trivial matter now. Data is a commodity. It is a commodity, which means you can buy whatever data you want pretty much whenever you want it to some degree. So, given that it's a commodity, you have all of these what would be protected factors in very rich data sets and they're being used for all sorts of decisioning purposes.

A good example of this is what I call proxy credit scores. They're not formal credit scores because they're not using the same kind of credit report data that is regulated, but they're using other factors that mirror that same data. And, so, let's say you've taken out all clear indicators of race or maybe even marital status, there are other inferred factors that will then be in the data that will -- or can be used to do exactly the same thing. So, you take out one and it's like a jack-in-the-box, another will step up. And this is how large data sets become really problematic for ensuring privacy and fairness, because you have all of these redundant factors again and again and again in the data.

And how we focus on correcting for that problem is very, very important because, right now, we're not, not in very many situations. There's not one global solution right now that corrects for that problem, because that is not regulated data. So, we've got to focus on that.

MS. GEORGE: Let me just piggyback a little bit on what Pam just said about the richness of the data set. I understand that, for some communities, their information may not be included appropriately in the data sets because of the way they use or don't use technology. Does anyone have thoughts on why that is and how it can be addressed? Nicol?

DR. TURNER-LEE: Well, actually, I was going to separate into that, that puddle there. You know, I think that's an interesting piece because we often think about, you know, in these conversations, for those of us that are entrenched in the telecom space, you know, broadband adoption here, data here, you know, broadband-enabled applications here and actually all these verticals cross, at some point, to give us a rich robust conversation and story, right, on how all these things interface. And I would say, given -- I'll give a shout-out to the Center for Data Innovation who Daniel -- I saw him here -- who published a paper I didn't get a chance to read, but I got a chance to read it over the weekend on data deserts, and I'm sure he'll talk about it later, but if you think about the disparities in broadband adoption, you have 30 million plus people that are offline that are not contributing in any way possible to this ecosystem. To a certain extent, you also have people who don't have, as my buddy John Horgan has mentioned, the level of digital readiness to actually go online and engage in a very participative way on the internet for, you know, noncommercial value versus commercial value, et cetera, you put all that together -- and, I mean, I was thinking about your comments, Stuart -- you might begin to see some segmented marketing to some of those folks because, you know, you have the others, the sociologists, the perspective that my online behavior may match what I do offline. And, so, I may be looking for, you know, something that I may not perceive to be predatory in the offline space translates to what I'm searching in the online space, which then leads to some type of predictive marketing in the types of products and services that I use.

So, I think we have to solve that problem. And I constantly tell people the broadband adoption digital divide issue has not gone away, because I think when you have the dearth of data particularly for vulnerable minority populations and data is driving certain decision-making and driving certain efficiencies, you then disadvantage a whole group of people that, in some way, to your first question, right, could benefit from the positives of big data. They get left out or their results get skewed because the proportion of people that are participating may not have these other factors that, you know, the literacy and the readiness at hand to equally participate.

So, I think the inclusion piece, you know, the Center for Data Innovation, just a last point, calls it the data divide, you know, it still goes back to the data and inclusion divide on how you look at this big picture.

MR. PRATT: So, I would add that one of the approaches our industry has taken, though, whether it's a fraud prevention tool -- and by the way, we live very much in the fraud prevention world and in the -- sort of the ability to pay world and really everything -- all that data that flows into that transaction, for example, where I've made an application. Of course, it's a question of what application am I making and when did I learn about it and those sorts of things as well.

But we sometimes look for -- I'm going to use a term that we've used at CDIA -- necessary services, so ubiquity. In other words, there is a question of that. In other words, when you pick new data sources and you're trying to use a new data source, you want a data source that is broadly used. And, so, utility data is, by example, a type of data because virtually anyone who has -- no matter where you live, you are likely paying for a utility of some sort. It could be very straightforward, you know, water service and this sort of thing, electricity, and then telecom is an example of, again, where you have a fairly ubiquitous set of data. You're pushed deeper into communities that are economically disadvantaged who may not actually be engaged in a lot of the other types of credit activities.

I serve on a World Bank task force. We talk a lot about this. In fact, we're flying in probably 30 central bankers to Dubai for a meeting to talk about data sets that can be used in various parts of the world to create SME-based lending, which is often, you know, small to medium enterprise lending, but it ties in with really personal loans as well. It's almost the same thing as conterminous in a lot of places. But the idea is what data sets are out there. Colombia, for example, not South Carolina, Colombia uses telecom data widely.

By the way, the Credit Builders Alliance is a great group to take a look at when it comes to trying to segment the population of consumers who may be credit invisible. So, for example, Credit Builders Alliance focuses not on the under-banked, but really on the unbanked, those consumers who probably have the greatest financial stress in their households. And there's a group called Axion down in San Antonio, Texas, and they're experimenting with different data systems, which are interactive with the consumer, to try to build a data set which allows them to predict success.

CBA aggregates these small loans that are urban-centered loans, that are often minority-focused loans, that are sometimes tribal-lending systems as well, and that data flows back into traditional credit reporting systems, for example. We have other members, for example, who aren't running a traditional credit bureau, but have stood up completely new data systems -- Mark discussed one of them -- where we can reach new populations for the first time using entirely different data systems that aren't just simply built off of a traditional credit report, that are built otherwise. And, in fact, I think five or six of our members, along with CBA, we sponsored a symposium on this earlier this year. It was hosted by Pew, but it was run by Credit Builders.

I think it's a pretty good intense dialogue and, obviously, you know, dialogues like this inform our thinking in terms of how we go forward and what are some of the framing issues. But I do think when you have an Equal Credit Opportunity Act, a Fair Housing Act, even universal service pressures that are put on the telecom industry, those drive industries to think about whether they have a Community Reinvestment Act obligation or not, it drives industries to think about how do I reach communities that are harder to reach otherwise and in what way.

Under-banked have different needs than unbanked, depending on definitions. Under-banked have different needs than middle class consumers, though, who are still living in very tight circumstances. And, so, as you move through societal tranches of consumers, the kind of data that we have allows us to work through that and to, again, match up a better offer, we hope, an offer which leads to success on both sides.

MS. GEORGE: Okay. So, I want to talk a little bit more about this notion of privacy, which some of you have touched on. And we've heard some mention in the comments to this workshop about the role of data-obscuring technologies or techniques or privacy-enhancing technologies, such as de-identification. Is there a role for those types of techniques going forward and are there some that are better than others? I know Cynthia wants to say something.

MS. DWORK: I think that privacy and fairness are completely unrelated and simply don't understand what de-identification would have to do with this discussion at all. But going back to privacy or questions of hiding information from the classifier, as Alessandro said, I do have some examples there.

So, if you have a really well-trained classifier and if you want to train a classifier well, you want to give it as much information as possible. So, for example, hearing voices may be diagnostic of schizophrenia in one population, and in another population, it might be part of a common religious experience.

You could have, theoretically, a minority group that is -- in which bright students are steered toward mathematics and you might have a majority population in which the bright students are steered toward finance, and if the minority is very small compared to the majority and you're looking for a quick and dirty classifier to find bright students, you might just look for finance. But that would be neither fair to the minority, nor would it be giving optimal utility because you would miss out on the gems in the minority.

And, so, there is a role for using as much information as possible, and withholding information would be inappropriate in those contexts.

MS. DIXON: Well, you know I've got to respond to that, right?

MS. DWORK: Go for it.

MS. DIXON: Okay. So, I do think privacy and fairness are aligned and very important in fundamental ways. But I think it is in ways that are actually surprising when you start to think about them at the deeper levels. So, let's look at large data sets and analytics in terms of, you know, the structures that can govern some of the new things that are happening. So, fair information practices -- well, wait, let me take a step back.

So, first off, I said earlier that big data is immature. It is. It is immature and there are two really big markers that tell me that it is an immature -- in an immature state. Number one, there is no firm scalpel-like legislative definition of big data. Now, I know what big data is, we all do in this room, right? But show me an actual legislative definition of it, and I know that you can't right now because there isn't one yet. There will be, but not yet.

So, the second thing that indicates that big data is currently a bit raw and unformed is there are no global solutions to the various problems that it poses. Right now, though, there are focused solutions and what I would call also local solutions to specific problems, surgical strike solutions, and there are also ways of -- so, those are the two things that exist. But how do we -- so, we're clearly at a formative stage. So, what do we do with that?

We can't just throw out the existing fairness structures. Some have said, oh, big data, okay, let's just push everything aside and let's start from scratch. I don't think that's necessary or appropriate at all. We need to use the existing fairness structures that we have, Equal Credit Opportunity Act, Fair Credit Reporting Act, HIPAA, the Common Rule, the Belmont Report, the Nuremberg Code. These are ethical codes, of course. And then, of course, the Fair Information Practice

Principles, these are very important. We can't just toss them out because there are some weird things happening. So, we need these old structures.

And on top of that, to address your question, what do we do, we need to look at what do we do in terms of what I would call statistical parity. We have to have statistical parity, statistical fairness. And there are ways of achieving that. So, it's these fairness structures and statistical parity.

So, for example, Stuart said something very compelling about how you're choosing the data sets. That is part of statistical parity. Where are you getting your data? Was it from people who volunteered this data or was it coerced? Was there mandatory classification of people? Was someone put in a box in a mandatory way that they maybe didn't want to be or didn't know about? So, these are all very significant considerations in how we deal with the fairness and privacy piece, because there is information that is so deeply prejudicial that it really is a classifier killer.

So, for example, if someone is found to have HIV/AIDS, it really breaks a lot of the classifications that they're in and really impacts the outputs. And in other language, that might be called sensitive information, but it's also highly prejudicial, and we need to really understand that privacy has a role in this because there is some information we need to think about not collecting, and if we do collect it, we have to protect it. HIPAA was right in how it handled that. It handles medical research for human subject research protection, there is very meaningful robust consent in what's called an IRB process, Institutional Review Board. And, so, there are examples already in place where we can go.

MS. DWORK: So, first of all, having worked for more than a decade on privacy preserving data analysis, I don't want anyone to think that I don't care about privacy. I do care about privacy. I'm just saying that, intellectually, mathematically, privacy and fairness are not necessarily the same thing. What you're talking about is the inability of the people who are making decisions to disassociate certain pieces of information from the decision. And what is really going on here is that you're searching for -- and very, very appropriately -- you're searching for some kind of a measurement for any particular classification test, you're searching for a way of measuring how similar or dissimilar are two people for this particular classification task.

MS. DIXON: That's right.

MS. DWORK: And quite possibly, the very best measurement that society and math together could come up with would involve all sorts of factors. But you don't trust the people or the machines or whatever that are making the decisions right now to give them all of the information, and that's probably very reasonable.

MR. MACCARTHY: So, let me jump in here. I think this -- you know, this is a very abstract and almost philosophical question. If you look at some of Cynthia's work, I was just telling her she defines this concept of relevant similarity as a way of first saying do that and then go into maximizing utility. We've heard that before. Immanuel Kant said that in his theory about ethics. So, we're dealing with some pretty abstract and philosophical questions when we come to this stuff.

And at the level of social policy, at the level of what we think is fair and what we think is just, I think a lot of the discussions we're having here, they may seem to be about data and how to interpret data and so on, but I think they really go back to some of these basic ethical and philosophical questions. So, I do think we need to take a step back and not to think about these issues as if they were issues about data and analytics, but they really are pretty broad social questions.

So, for example, do we need to have a special social policy towards big data? My instinct is no, big data is just an evolution of what’s been going on in the data analytics world for generations and to think we need to have a special set of laws or best practices just to pick up the big data subset of all data analysis, I think is the wrong direction to be thinking about. I do think we need to focus not on kind of global solutions to all these problems, but to go back to the specifics.

As Stuart’s been saying, you know, there is a well-developed body of law that surrounds certain uses of information and we’ve chosen to put that body of law in place because we think, in those areas, concerns about social policy are the greatest and, so, we need a large sort of set of protections for that.

In other areas, where Mallory was talking about sending catalogs to men rather than to women or advertisements for cars that appeal to men, our social concerns are a whole lot less. So, the idea that we would have one set of rules, one set of fairness requirements, one set of access requirements that goes across all data uses, I do think that’s the wrong direction to go in.

DR. TURNER-LEE: So, I want to jump in because I think I agree, to a certain extent, though, with regards to having some framework, though, of what transparency and the purpose of your data looks like. I mean, I’m a big fan of the FIPPS, to a certain extent, when it comes to privacy concerns, because I think that people have to understand that their data is being used for particular purposes.

And in the internet, while I agree with Stuart that you actually have different bodies of policy buckets and privacy parameters that actually define how your data is being used, let’s face it, the internet is this big, big buffet of places that you can go. It’s not that simple any more to actually say, well, I’m going to the internet for this or I’m going for that. You know, people are going to the internet to engage in a multiple range of activities that, at some point, get muddled because it’s not necessarily going into your Safeway and giving your email address so that you can get benefits on your grocery shopping at Safeway, right?

When you give your email address on the internet, you know, there’s a data information service that is taking that information and creating algorithms of where to direct you and how to advertise towards you. There’s probably a search that you did that brings up, you know, a healthcare provider. You know, you might have gone and bought red shoes and the next thing you know you’re getting red shoes advertisements, ladies, right, for just one purchase that you made.

So, I think it’s such a hard ecosystem to sort of distinguish between this is why people are going to the internet for this particular purpose. So, I think a general framework, like the FIPPS, is actually appropriate to help us figure out how do you ensure that the input of data, whether it’s big or, you know, small data, does not impute cultural stereotypes as well as cultural cliches that actually lend itself to predatory behavior and actions on the part of, you know, the online space. I think that’s so important.

I mean, we’ve seen it with segmented marketing where, again -- you know, again, for people of color -- and this is interesting because I’m doing a paper on this -- from the long term, we’ve not been able to see the exact civil rights infraction that happens because, you know, someone has seen something on my Facebook page or I put up a post. But it’s going to happen. It’s just a matter of time that we’re going to see that type of predictive analytics or algorithms defined and, you know, discriminate against people.

The question becomes, do most consumers know that when they participate -- particularly for minority consumers who over-index in social media when they are on and over-index, you know, on the internet as new users because they're experimenting, exploring and trying to attain the aspirations of other internet users, do they understand how their data is being used? Do they understand what distinguishes their private personal identifiable data from data that they're actually basically contributing to the ecosystem, you know, just because they want to be part of the conversation?

And, so, I think those are clear distinctions. Again, it was brought up in your paper, Alessandro, about that. But those are things that we look at at MMTC, you know, will that have an impact on someone's ability to get a job or healthcare or, you know, something of social value, not necessarily their ability to stream content, but something of social value that will essentially -- you know, when they are applying for a car loan, you know, will give them higher rates, and I think that's really important to put in this conversation.

MR. ACQUISTI: I wanted to connect what Nicol just said to something Cynthia said and something Solon this morning was mentioning. So, I'm ready to believe that most of the times more data may decrease discrimination, increase fairness, increase efficiency, but it's also the case that the opposite may happen. Some examples were given this morning by Solon talking about when data mining discriminates, and the other point was made by Cynthia when it is the human decision-maker with his heuristics and biases, which makes incorrect or biased usage of the information or even analysis made available to him.

The case in point Nicol was referring to was this experiment we did on the impact that social media information has on the hiring behavior of a U.S. employer. So, we did this experiment in which we applied to over 4,000 American employers, we have CVs, resumes, which were identical in terms of educational attainments and professional achievements for different candidates. However, we had also created social media profiles for these candidates. So, we wanted to see whether employers would go online and search for the personal information.

And employers did. And what was interesting is that they would react to the personal information, specifically to disclosure of a religion affiliation, in a discriminatory manner so that our Muslim candidate was less likely to be invited for an interview than our Christian candidate, and this is a parity of professional and educational background. So, this addressed a potential problem sometimes with more information not necessarily leading to more fairness.

There is also a broader story, which is the huge tension that this kind of study shows between the legislature who decided to have regulatory protections to certain traits so that certain traits should not be asked about in interviews or should not be used in the hiring process and information, not just information technology, which is effectively bypassing the legislation because it’s making this new data, these traits, these attributes perfectly easily available to employers without employers even needing to ask during an interview.

MS. DIXON: You know, there’s a really interesting idea here and I want to jump into the weeds a little bit to explain it. So, earlier in my comments I talked about the fact that when a person is classified, it triggers the data paradox. And really we could spend many hours talking about good big data and bad big data. All examples exist from the top to the bottom of the spectrum. We can take that as a fact and just move forward with that.

And then here’s the deal though. So, in regards to your comments, Nicol, I was, you know -- one of the difficult things that I was forced to unambiguously assent to at the conclusion of the researching of the scoring paper is that really we cannot control our information flows any more, our so-called digital exhaust. We really don’t have the full rights and tools to shape them right now. And one of the really big ways this is happening is in retail transactions.

So, if you look at a lot of the data broker lists and a lot of other data about how our data is being gathered for classification, one of the big ways this is happening is through the analysis of our retail purchases, and it’s like, okay, so who’s doing this? Is this just, you know, debit and credit card? How is this happening and can I opt out? Is there a notice about this? I think this is a very in-the-weeds specific example of you don’t have to be on social media to have this issue impact your life. And we’re talking about long-term, you know, big patterns here. You know, is someone purchasing over-the-counter medication? Is someone purchasing wound care for someone who had a serious injury? Is someone a diabetic because they bought a magazine, you know, that may infer that?

And then we can game it on the other side. Did you buy hiking boots? Did you go to REI? Are you subscribing to a running magazine? Cool. This will help your -- perhaps your health plan to charge you less. So, you can game it on all sides.

But the question we really have to ask going forward is what’s happening here and what structures can we use to ensure that there is fair information principles that are encoded into all of these processes from top to bottom, so that when we make a purchase, we're confident that what we're buying, we can use our credit cards, we can use our debit cards. We don't have to run around like some crazed tin-foil hat person and use cash for everything. That's not the answer. The answer is fairness structures that protect our digital exhaust and that give us the tools and abilities to shape it.

I've actually been heartened by some of the opt-out tools that I'm seeing that are pretty granular and that allow us to see where we've been categorized and then choose and alter our categorization. This is very helpful. So, Acxiom has one of these. Their opt-out -- their data -- about the data portal. I went and looked at my categories. I have very different categories depending on which email address I use. And, so, I did some granular opt-outs and feel much better about the world. Now, I won't be seeing advertising for Asian men. Someone thought I was an Asian man. I don't know how they did that, but anyhow. So, categorization is a big deal and it can really change how your life looks.

MS. DWORK: So, I'd really like to bring up a paper here that just floated across my desk, and I'm afraid I don't even remember the entire author set. Anupam Datta was one of the authors. But it was a -- the paper involved experiments that were done in which people had changed their categorizations on Google and it did not have the anticipated change in advertising.

MS. DIXON: Oh, interesting.

MS. DWORK: So, I’m sorry I’m not informed in more detail, but I suggest that people look this up.

MS. DIXON: Yes. That’s interesting.

MR. PRATT: You can see how in this dialogue we’re beginning to sort of categorize uses as well. In other words, categorize -- and I think that’s important that we begin to unpack this dialogue and not allow big data to just get squished together into a sort of singular dialogue. The kind of data sets that a CDIA member has are really -- they’re not often -- and certainly not for risk management purposes, kind of big data that is derived from my search engine searches, the websites to which I go.

There are some lenders that are experimenting with the use of that kind of data. Consumers are essentially opting in to do business with that lender. It is important to know that that lender is still obligated to live by the Equal Credit Opportunity Act. So, even though -- so, there’s an example of a lender with kind of a closed system of data and the consumer said, yes, you can use this data. I don’t have traditional data sets, you know, for you to be able to make that lending decision. So, I do think that's occurring.

Also, we haven't talked too much about it and I'm not sure that these terms apply quite as often today, but really structured versus unstructured data is also part of the discussion. You know, unstructured data might be data that's more so less directly identified with me. It depends on whether you think an IP address is personally identifiable information or not.

MS. DWORK: Yes, it is.

MR. PRATT: No, it's not.

MS. DWORK: Yes, it is.

(Laughter.)

MR. PRATT: And later, we're going to be doing a little song and dance, it's going to be really good. But I would argue that IPs can be associated with individuals. But the question is, our databases that our members build are still based on identifying information of the traditional type because our members are building -- if they're building a database for purposes of an eligibility decision under the Fair Credit Reporting Act, then they have to build the database along a certain set of lines to make sure it's accurate and meets the accuracy standard. And this kind of goes to the point.

So, one of the questions is whether you use the FCRA as the template or whether you use a fair information practices template of some sort, and there's many of them out there, I tend to like APEC'S better than some others -- you know, the question is when do you apply the template and in how nuanced a way do you apply that template to that kind of information.

So, there's a lot of advertising activity going on out there. Our members -- like I said, our members tend to have a structured data set. It tends to be built off of identifying information. It tends to be wrapped in a law, like the Gramm-Leach-Bliley Act. You can build a fraud prevention tool to protect consumers, but it's not going to stop a transaction, it slows it down.

Essentially, it's like going through the metal detector and then having somebody wand you to make sure that they really know whether or not you're carrying something into the building versus eligibility. I want to get into the building and I need to have a certain set of credentials to get into that building and can I have access to those credentials and how are they used and so on.

We're a very use-based society, by the way. We look at outcomes and we tend to measure data uses in terms of the outcome, as opposed to trying to manage each step of the process. I had a -- I was on a panel in Berlin where, oddly enough, milk production was used as the example here in terms of regulatory strategy. And at least in Germany, this fellow, this economist described the German government regulates every step of the process in milk production. So, really it's a -- forgive the pun -- a homogenized approach to milk production. You really have no strategy by which you're going to be able to remove cost from the market and be able to improve your margin even if you have a very -- you know, a very structured price structure on the back end.

Here in the United States, we don't tend to regulate every step of the milk production process; we test at the end to see if the milk is homogenized properly, if it is -- meets the purification standards and so on and so forth.

So, we're kind of getting deep into this very, I think, almost philosophical discussion, as Mark termed it, and I think that's right. What template do we use for what type of use? When is categorization an issue of harm, for example, might be one way to think of it. When is categorization just a question of whether I got a catalog that was applicable to me as a buyer of certain products in the marketplace?

But I do think we're doing pretty well as a country in terms of eligibility. When data is used as a gatekeeper, that data is regulated by a fair information practices structure under the Fair Credit Reporting Act. When data is used for fraud prevention, there's a law that wraps around it. When data is used in all those transactions, there's quite frequently, in fact, very definitively in the context of insurance and in the context of credit and fair housing, in particular, and the equal -- and then the EEOC as well, there are laws which establish the baseline result that we expect, and we expect to see a result which is fair for all, fair treatment for all, and that we've even established, rightly so, protected classes, because we have found problems in our society where we did not identify these protected classes.

MS. GEORGE: So, that's actually the perfect segue to my next question, which was, as we move forward in this era of big data and these new practices, what is the model? Should it be based on use? Should it be based on harm? Should it be based on data collection methods, active versus passive? Like what are the guideposts that we should be looking for as we emerge into the future?

MR. MACCARTHY: So, let me quickly jump, if I could --

MS. DIXON: I'll go next.

MR. MACCARTHY: Yes. I think you touched on the two big ones, which are use and harm. This brings us back to the, you know, very specific discussion of very specific ways in which information is used and how people can be damaged. And I do think we -- sometimes more information is better in order to achieve the particular outcome that we want. Sometimes more information is not so good. I mean, there's the famous experiment, natural experiment in why classical orchestras were all men for years and years and years. It was because the conductor would look at the people who are actually performing the music and notice which ones were men and which ones were women. But when you put them behind a barrier so you couldn't tell what the sex was, suddenly, it became 50/50, you know. Withdrawing information, in that particular situation, was something that was very helpful in avoiding a discriminatory problem.

For many uses of racial and ethnic information, the decision makers aren't even allowed to know about race and ethnicity. So, we want to keep that information secret. Maybe privacy there promotes fairness.

But sometimes more information is more. All these products that we've been talking about, the alternative data products, they require more information about people in order to accomplish their good purpose.

The -- another example, and this goes back to your point, will businesses and others, you know, try to reach out and try to solve these problems? Well, most companies want to have a diversity program where they reach out to make sure that their workforce looks like America and they want help to do it. There's a new service provided by a company called Entelo that will use information, social network information, information on the web, it's in the FPF study, which I mentioned earlier before. And the idea is using this kind of proprietary tool, you'll be able, as a company -- be able to target your recruitment efforts to try to get at the kind of people who will be qualified for your work and yet will satisfy your diversity requirements.

So, the uses of information, how much you need, where it comes from, how it's used, those are all relevant factors. I don't think there's a template, there's no one-size-fits-all, here's how we do it all circumstances and for all purposes. But I do think if we pay close attention to the actual uses and the dangers we're trying to guard against, we can make some progress.

MS. DIXON: So, great question, and I appreciate your comments, Mark. They were very thoughtful.

So, I want to talk about medical just really briefly because it really does provide a really intriguing example. So, if you look at the issue of medical research, a lot of folks will cite medical research as a perfect example of how to handle big data. And, you know, medical research is intriguing on a lot of levels. If you look at the various ways that the ethics of how privacy works in the medical field are crafted, it's absolutely fascinating.

So, to kind of dive in, if you look at human research subject protection, that's where the strongest medical privacy protections are, if you're doing research that impacts human subjects. So, if you're federally funded, you're going to be captured under something called the Common Rule. The Common Rule is a regulation, so that is regulated. You will have to get meaningful consent from the individual in order to participate, and it's all run under an IRB process.

That Common Rule is very complex and it was built on something called the Belmont Report, which was not a piece of legislation. The Belmont Report was built on something called the Nuremberg Code, which was an ethical code developed after the World War to prevent any kind of human research atrocities from ever occurring again. The Nuremberg Code had, as its absolute bedrock foundation, human consent as absolutely the bedrock of what has to happen in human subject research protection.

And even though the Nuremberg Code was an ethical framework that didn't have legislative teeth, the teeth it had is that it appealed to our humanity, and that's what stuck. It stuck all the way through the Belmont Report, it stuck all the way through the Common Rule. And where we see it violated today, in certain commercial instances, it strikes us, again, as an unfairness.

So, it's very important that the ethical frameworks are also considered in adjunct and in addition to the regulatory frameworks that exist because they all have something to add. And in cases where regulatory frameworks do not apply because of narrow applicability, we really need to look to the ethical standards because they are human. They say something human about us and it's what's really important to listen to.

MR. ACQUISTI: You were asking about what model may work. I am on record as criticizing transparency and control mechanisms due to a series of behavior experiments we have run showing how, for instance, control of personal data or even just a feeling of control of personal data can lead to more risky disclosures, over confidence, and more risky disclosures and transparency is very ineffective in that I can read something, understand it, and then that information is no longer salient at the moment I have to make an actual decision.

However, let me for once actually take the defense of transparency, in fact push even the envelope farther, kind of maybe a little provocation for the panel, and focus on the concept of data provenance. What if we start applying the rules of the data industry once we use it on consumer data, we apply the same rules from the consumers on the data that firms have about consumers.

So, imagine a system where we -- every piece of personal information held by any data holder has to be attached to metadata showing the exact provenance of that information, whether it is observational data, data traded and received from another entity, or inferred data. So, data predicted based on some algorithm, in which case, also, the algorithm should be revealed. If I am classified as a consumer who is willing to pay $80 for this good rather than $40 for this good, I would like to know why.

Considering the sophistication of the data, the way it is presented to us, as nearly being able to solve, in the close future, any societal problem, that kind of technology, of attached metadata showing the provenance of personal information is not really that far -- that science-fiction-like. Otherwise, if you keep having big data for consumers and only trade secrets for firms and how firms use data, that's the kind of information asymmetry which economic literature tells us will reiterate rent positions and economic imbalances.

MS. GEORGE: We're drawing to a close here, so I just want to remind the audience, if you have any questions that you'd like to submit to the panel, we have staff around the room who can collect your question cards. And in the meantime, I'm going to pose one final question to the panel before we start wrapping up.

So, on this notion of transparency and control, there's been some suggestion that providing more control to consumers is the solution to the problems of big data, providing technology and techniques for consumers to be able to control how their data is collected and what happens to them. Are there limitations to that proposal or is that the solution to this problem that we've been discussing? And, Nicol, I want to start with you.

DR. TURNER-LEE: Yeah, I mean, this is a very interesting question because this whole time I've been talking about empowering consumers, right. But I think it was mentioned earlier about this whole concept of opt-out, right. And because there's going to be some data that we need that have socially beneficial purposes, that we would like most people to participate, energy being one of them or any type of utility. We would certainly want people to partake in it because it's a passive data collection, not necessarily an active data collection because we're essentially gathering information about the utility use that will prove valuable to us in improving, for example, the smart grid or other things in our society.

At the same token, and this is a conversation -- I was joined by several scholars, on the internet of things, you know, when a person, for example, walks into a home that is fully wired because of the internet of things, your toaster, your refrigerator, your bed for that matter, all registers personal data, do you have the ability to opt out of that environment just because you don't want, you know, people to see how often -- you know, if you're like me, you don't make it to your bed often because you're also reading papers and you're sitting on your couch, right.

So, it's like, you know, at some point, I think the conversation has to be made and I think we've all touched on in some way to your earlier question, Tiffany, about, you know, when we're coming up with a framework, does it balance use versus harm, right, with allowing some flexibility for the collection of data that will help us for the purposes, again, of efficiency and public good, and the extent to which consumers, you know, from the front -- I mean, that’s another -- when we start talking about this -- and, you know, not to make this long-winded, but when we start talking about this, when I was at Joint Center for Political and Economic Studies years ago, we did just a raw review of privacy policies and we recognized that, in some cases, you had to have a PhD or a JD just to read the privacy policy. You know, after we ran them through the fluency indicator, you know, the level of what people are engaged in is sometimes not known, you know, in terms of what they’re actually getting into.

So, I think the opportunity to look at creative solutions, like an opt-out or allowing people -- you know, we should not have it where we look at consumer protection when a bad actor, you know, comes to the play or a bad action happens, because that’s probably hardest to actually reverse at that time, particularly for, again, minority communities. When your credit is compromised and you don’t own a home or you don’t have a bank account, the biggest asset you have is your social security number. Imagine what it’s like for a senior

African American woman to have to repair her social security and her credit, you know, because of an infraction of harm.

So, we have to figure out ways for people to have a lot more knowledge as to, you know, one, the internet is a participatory environment and, in some cases, you'll know when your data's being collected and sometimes you won't, right. Two, when I feel that there is some particular harm or some type of compromise in terms of my personally identifiable data, in particular, right, I have that decision to opt out. And, three, going back to my earlier notion about the internet of things, I have the ability to say I don't want my data looked at if it's pertinent to me as an individual, you know, and not necessarily something that's more pertinent to the broader group.

So, I'll pass it over to you.

MR. PRATT: Thank you. Yeah, I think it's, in some ways, an all-of-the-above strategy, meaning you really need to look situationally at the nature of the data and really fair information practices are not a model --

DR. TURNER-LEE: Right.

MR. PRATT: Even if you were to look at a FIPPS model, it's not monolithic. I remember working with the

GAO group a while ago. They were looking at government uses of data and they applied an OECD FIPPS model, but they did it in a really clumsy and sloppy way, and it was really rigid and it didn't make a lot of sense. But I think having framework models to trigger thinking and create more sophisticated analyses and understanding is very important, and I think a number of the academics in this discussion already have introduced papers, as well as thoughts, that suggest that data which seemingly is neutral may not always be neutral or an algorithm which we think is neutral may not always be neutral. We should think about that, and that's part of our FIPPS model, if you will. It makes a lot of sense.

But opt-out will work in some cases and opt-out won't in others. A great example is years ago I remember there was a -- one of the browsers had given me the option of turning on a switch, if you will, so that I could track cookies and I could decide which cookie I wanted to accept and which one I didn't, except that every time I went out onto the internet, my screen was just covered with little cookie notices and it was almost like pop-up ads. I mean, I was clicking and clicking and clicking trying to get rid of all the -- you know, the damn cookie notices. And before you knew it, I was not reading the cookie notices; I was just doing battle with them, right, to kind of -- so I could actually see what was on the screen.

So, there would be almost like a behavioral issue there for consumers, right. You know, how do consumers behave and what is the -- what is your goal and what's the most effective strategy to kind of get to that goal. So, I'd say it's kind of all of the above and it's nuanced and it's careful and it's thoughtful and it's probative and it's not just simply this monolithic -- which is what I think is sometimes the problem with law. Law often is too monolithic and too rigid and is applied in a very sloppy way and it can be harmful.

A great example would be HMDA data, Home Mortgage Disclosure Act data. If we're trying to determine whether or not creditors are -- even if creditors themselves are trying to determine whether or not they have a practice which is facially neutral, but is not in some fashion, it's hard to know that if you're not gathering the data set that you need in order to then look for that in order to decide, wow, okay, I have something here that I couldn't discover in the first place because I'm prohibited from gathering the racial information that I might otherwise need. That's the nuance of it, I think.

MS. GEORGE: So, I see we only have a few minutes, so I'm going to ask if anyone has any final thoughts because we don't want to keep people from their lunch.

MR. MACCARTHY: The only quick thought I've got is that this focus on use and harm is a really alternative way of thinking about these things. If you put too much weight on the alternative of giving information to users, being transparent and then letting them choose, that's really your focus and you're really pushing that as your major defense against unfairness and privacy invasions, you got to do it. In some cases, human subject experimentation is not something we want to sort of make decisions for people. But if that's your universal solution, I think you're really doing customers and consumers a disservice.

You're responsibilizing your own users, you're telling them it's their problem, you figure it out. Here's a bunch of data you don't know anything about or how to interpret it, but I've given it to you and if you want to opt out, go ahead, opt out. I think that's not a productive way to protect people because the tendency for people in that circumstance will simply be to throw up their hands and do something else.

And on the other hand, if you make the person who's gathering the data and using the data responsible for fair and appropriate use, that I think points in the direction of putting the responsibility more where it lies, not simply on the data subject to protect himself completely.

MS. DWORK: So, that actually comes back to the point that I made at the very beginning. I think everybody should be thinking all the time about, for various kinds of classifications tasks, who should be treated similarly to whom. And we have got to start, as a community, taking responsibility for trying to lay out those rules. This was done in the context of fair credit reporting; it should be done in lots of other contexts as well.

MS. DIXON: I don’t think the structures need to be reinvented or shoved aside because data sets are larger. It’s important to keep the regulations that we have, allow them to apply where they’re applying, to ensure that fair information principles are applicable and still relevant and still practiced, and we also need to add statistical parity and we need to look at the underlying ethics of the issues as well, because where there are not frameworks, there still are underlying ethics and we can’t ignore them because some of the problems that exist in the uses of this data are fairly profound and there’s a lot of discussion of, oh, well, let’s -- you know, let’s throw out collection limitation because it’s too hard and let’s just focus on uses. And then there’s discussion of, oh, well, let’s not -- let’s not control uses, let’s focus only on collection limitation.

Look, right now, we’re in a situation where we have many multiple overlapping remedies and I think that’s going to be the case for quite some time and we need to look at those remedies, really study them, see where they’re working and how, and look to see what’s important and what we need to focus on, where are the real problems and where are the most disparities occurring, and let’s fix those and move through the ecosystem with it.

MS. GEORGE: Alessandro, you have anything else?

MR. ACQUISTI: In essence, my final remark was my point about the provenance, data provenance, and kind of applying the same rules of big data to consumers to firms’ handling of consumers’ data.

MS. GEORGE: Well, thank you very much for this lively discussion. We did get a couple of questions at the end which we’re not going to get a chance to discuss, but our panelists, I think, will be around this afternoon if you want to talk to them. I want to thank each of you for attending and enjoy your lunch. I hope you join us for the afternoon where we'll begin with a lovely presentation by Latanya Sweeney. And thanks again to each of our panelists for presenting.

(Applause.)

(Whereupon, a lunch recess was taken.)

AFTERNOON SESSION

(1:22 p.m.)

MS. ARMSTRONG: I think we're going to get started in a few minutes, everyone.

All right, gang, how was lunch? Okay, time to find your spot without your food or beverage.

Great. Thank you, everyone for joining us this afternoon. The afternoon session is going to be -- Commissioner Brill is going to give opening remarks to the afternoon session and so without further ado, here is Commissioner Brill, who needs no introduction.

(Applause.)

REMARKS BY COMMISSIONER JULIE BRILL

COMMISSIONER BRILL: Thanks, everybody. Before I begin, let me just say thank you so much to Katherine, to Tiffany, to Patrick Eagan-Van Meter
, and Katherine Worthman,
to all the folks at the FTC who have been working so hard on this workshop. I think that the quality of the panels this morning, the quality of the panels this afternoon, show you how much work they put in to organizing this event. So, can we just have a quick round of applause for the FTC staff.

(Applause.)

COMMISSIONER BRILL: And thanks to all of you who are watching by webcast and those of you who made it here today.

Our presenters and panelists are providing us with details about the current and emerging uses of big data to categorize consumers, the surrounding legal issues and possible best practices for big data analytics providers.

I'd like to provide a more general, and also, perhaps, a more personal perspective that makes, I hope, a simple point. Providing transparency into big data algorithms that categorize consumers has been done before. It has put some concerns to rest and companies and consumers have been better off as a result.

Now, as I've said on one or two other occasions, those of you who have read some of my speeches or perhaps attended them, I believe big data analytics can bring significant benefits to consumers and to society. But we must endow the big data ecosystem with appropriate privacy and data security protections in order to achieve these benefits.

Today I'd like to focus on three of the more challenging issues of the intersection of big data and consumer protections that pertain to this workshop. I'd also like to offer some suggestions, some specific suggestions, about what industry can do right now to address these concerns.

Consumer trust is critical here and transparency and accountability are key to building it. Now, the first challenge involves traditional credit scores derived from credit reports and alternative scoring models. In this realm as in many others, past is prologue.

The origins of the Fair Credit Reporting Act have something to teach us about our current environment. The FCRA was our nation's first big data law. The seeds for it were planted in the growing economy after World War II. Businesses formed cooperatives to enable quicker and more accurate decisions about creditworthiness by sharing information about consumers who were in default or delinquent on loans. Over time these agencies combined, paving the way for consumers to again access to credit, insurance and jobs.

As credit bureaus increased their ability to draw inferences and make correlations through ever-larger databases, unease about the amount of information the credit bureaus held, as well as its accuracy and use, also increased. Congress passed the Fair Credit Reporting Act in 1970 to address these concerns.

The FCRA governs the use of information to make decisions about consumer credit, insurance, employment, housing and other transactions initiated by consumers. It covers not only credit bureaus, but also, importantly, their sources and their clients.

The FCRA gives consumers important rights. For instance, consumers are entitled to have access to their data, to challenge its accuracy, to have irrelevant data removed. And to be notified when they are denied credit or get a loan at less than favorable rates because of negative information in their files.

The use of credit scores has thrived under the FCRA's rights of notice, access, correction, relevancy and accuracy. And the FCRA has enabled the credit reporting enterprise to serve a purpose useful not only to the credit reporting agencies and their clients, but also to consumers.

The credit scores that first emerged from analysis of consumers' credit files broadened access to credit and made determinations of a particular consumer's worthiness more efficient and more objective, than the case was with prior, more subjective, determinations.

Now, as scoring models began to proliferate and enter into new types of decisions, including employment, insurance and mortgage lending, consumers and regulators grew concerned about what exactly was going on within these models. Some of the most important questions were whether credit-related scores were using variables that act as proxies for race, ethnicity, age and other protected categories.

In 2003 Congress directed the Federal Trade Commission and the Federal Reserve to study these questions in the context of credit-based insurance scores and traditional credit scores. After extensive and rigorous studies, both agencies found that the scores they examined largely did not serve as proxies for race or ethnicity. The FTC and Federal Reserve reports shed a lot of light on traditional credit scores and assuaged some important concerns, which was good for everyone involved -- consumers, credit bureaus, and credit score users.

Now, let’s fast forward to today. We’re now seeing a proliferation of other types of scores being used to make FCRA covered eligibility determinations. While these scores are subject -- or, many of them are subject to the same obligations of access, accuracy, security and other requirements imposed by the FCRA, they haven’t yet been subject to the same kind of scrutiny that Congress and the Federal agencies brought to bear on traditional credit scores.

The use of new sources of information, including information that goes beyond traditional credit files to score consumers, raises fresh questions about whether these alternate scores may have disparate impacts along racial, ethnic or other lines that the law protects or that should be addressed.

Those questions are likely to linger and grow more urgent unless and until the companies that develop these alternate scores go further to demonstrate that their models do not contain racial, ethnic, or other prohibited biases. These companies may learn that their models have unforseen inappropriate impacts on certain populations. Or they might simply find their algorithms should eliminate or demote the importance of certain type of data, because their predictive value is questionable, as FICO recently discovered with respect to paid off collection agency accounts and medical collections.

Just as we did a decade ago, the FTC and other appropriate Federal agencies should once again devote serious resources to studying the real world impact of alternate scoring models.

But industries shouldn't wait for Federal agencies or for Congress, for that matter, to get involved to review their own scoring models. Companies can begin this work right now and provide us all with greater insight into, and greater assurances about, their models.

The second big data challenge I'd like to discuss comes from the unregulated world of data brokers. As outlined in the Commission's recent report, as was discussed this morning, data brokers' profiles combine massive amounts of data from online and offline sources into profiles about nearly all of us. Data brokers' clients use these profiles for purposes that range from marketing to helping companies determine whether, and on what terms, they should do business with us as individual consumers.

Now, the main data broker issue that I'd like to highlight today concerns data broker segments that track sensitive characteristics, including race, religion, ethnicity, sexual orientation, income, children, and health conditions.

As I noted, when the FTC released its landmark report on data brokers, I see a clear potential for these profiles, ethnic second city struggler or urban scrambler, to harm low-income and other vulnerable consumers.

In an ideal world, a data broker's products that identify consumers who traditionally have been under-served by the banking community can be used to help make these consumers aware of useful opportunities for credit and other services.

However, these same products could be used to make these consumers more vulnerable to high interest payday loans and other products that might lead to further economic distress.

It all depends on how these products are actually used. Importantly, our recent data broker report did not attempt to analyze the harms that could potentially come from the uses of consumer segmentation of poor or minority communities.

Now, one of the reasons I support legislation to create greater transparency and accountability for data brokers, as well as their sources and customers, is so we can all begin to understand how these profiles are being used, in fact, and whether and under what circumstances they are harming vulnerable populations.

In the meantime, the data broker industry should take stronger pro-active steps right now to address the potential impact of their products that profile consumers by race, ethnicity, or other sensitive characteristics or that are proxies for such sensitive classifications.

Here’s what I’d like to see data brokers do. They should find out how their clients are using these products. They should tell the rest of us what they learn about their actual uses. They should take steps to insure any inappropriate uses cease immediately and they should develop systems to protect against such inappropriate uses in the future.

Now, the third challenge I want to mention relates to companies that use their own data and analyze their own data about their customers.

Companies, understandably, are eager to determine what makes their customers happy and how they can more efficiently service these customers. As they dive into their own treasure trove of customer data in order to offer perks or better deals to loyal customers, companies may also find that these common practices disadvantage certain groups of individuals, thereby, in the words of the White House's big data recent report, exacerbating existing socio-economic disparities.

Back in January, the Harvard Business Review asked companies to think deeply about where value-added personalization and segmentation ends and harmful discrimination begins.

Now, I want to emphasize that all of these industry players, traditional credit reporting agencies and their newfangled progeny using alternate scoring models, data brokers and the companies that use their products, and companies engaged in analysis of their own customer data, all of these players can take steps right now to address concerns about the potential discriminatory impact of their use of algorithms.

I'm hopeful that the same reservoirs of data that create the concerns I outlined will also lead to ways to get them under control. I encourage all members of industry to look for ways that the data in their hands could be used to identify disparate treatment along racial, ethnic, gender or other inappropriate lines, and to correct such treatment to the extent it exists.

Thank you very much.

(Applause.)

MS. ARMSTRONG: Thank you very much, Commissioner Brill.

Now, the next part of our afternoon agenda, before we get to the next panel, is going to be a presentation, Digging Into the Data, and I'd like to introduce LaTanya Sweeney, who's been the Chief Technologist at the FTC, and Jinyan Zang, a research fellow in technology and data governance. So, I'll leave you with the clicker.

(Applause.)

PRESENTATION: DIGGING INTO THE DATA

MS. SWEENEY: So it's great to be here. My name got mentioned a couple of times, so I feel like I don't need any other introduction. But I do want to thank Tiffany and Katherine and Katherine and Patrick and Maneesha and DPIP for organizing this and for allowing us this opportunity to present our work. Assuming I can get the clicker to work, because after all I'm the technologist, right?

So one of the things I wanted to also let you know is we started a summer research program under the guidance and leadership of Chairwoman Ramirez, who you met this morning. The idea was to bring in some of the best and brightest students and have them do research during the summer on areas of interest to the FTC.

Today we're going to report on one such project, but let me -- and we worked as a team, so all of the fellows were pretty -- kind of contributed to all of the efforts, but Jin and I primarily did the one that we're going to talk about today.

Krysta and Jim couldn't be here, but Paul is here, I'll just have him stand up. And the work that's coming out from the other fellows will be coming over the next weeks.

So, the Pittsburgh Courier was once the country's most widely-circulated black newspaper, it had a circulation of about 200,000. If you worked for the Courier or if you were to interview their staff back in 1911, they would say that -- your clicker doesn't work --

they would say that when an ad appeared in their newspaper they would review that ad. They had to review that ad because they didn't want to run the risk of alienating, isolating, or insulting the audience that they served.

Today, the Pittsburgh Courier -- the clicker still doesn't work -- is an online website and their ads are delivered through an online network for which no staff member actually reviews the ad. Instead, it's a big data analytic engine that delivers their ad.

Now, we all know the promise and we've heard a lot about the advantages of big data analytics and online advertising is no exception. It's not that you want just any old ad showing up anywhere, you want the ads organized so that the fisherman sees the fisherman ads and the young mother sees the baby products. And so that's the promise.

But in order to deliver that, there's a lot that happened to get that Macy's ad on that Pittsburgh Courier page. There are a lot of parties and a lot of different ways that can happen. So let me sort of just blow it up and introduce some of the ways.

So there's groups that will help you put together your ad campaign and your ad copy, help you find platforms on which to sell it. There are data brokers that are involved in taking the outside data -- is it the battery or is just I don't know how to push the button?

(Laughter.)

MS. SWEENEY: Data brokers taking outside data, bringing it into the online network, figuring out what it is to offer or what kind of offer, which ad would be the right one to target directly to you, and make that connection from end to end through that kind of network. And so that's normal and called targeted advertising.

But we're not going to talk about targeted advertising right now. Let's talk about something simpler, where it's only one party that's going to go from end to end, such as the Google network.

Google delivers more than 30 billion ads a day. And every ad is delivered in the time it takes to load a web page. That's -- I'm a computer scientist, that is awesome. That's really awesome. How do they do this? Well, we're not going to get into the specifics and I'm not sure everyone actually knows the specifics outside of Google, but we do know that there are billions of ads on one side. And what an ad bid is, is basically the copy, the ad copy, the key words of the audience that they would like to show that ad to, and how much money they'll pay either to get that ad put in front of the audience or for someone to click on it.

On the other side are these publishers, who will basically take an ad. And so Google gets to make the decision as to which ad is going to show up when.

We're very interested in how Google goes about doing that. Not so much about ripping open that cloud, that blue cloud, but understanding what effects might be on the outside.

So one of the things we did was we turned to Mixrank. Mixrank is a service whose whole business is about capturing online ads. So they survey the internet constantly, record every ad they encounter, where they encountered it, the data that was encountered, and so then you can look at the data through the eyes of the publishing side or through the advertisers. So this is an example.

One of the things they do is they get rid of behavioral effects and re-targeting effects. So this is nice for our study, because now we're looking at it with the assumption that that blue cloud doesn't know anything more about you than it would know about anyone else. And in those circumstances, how does the blue cloud perform.

So we found this website, Omega Psi Phi. Now, Omega Psi Phi had its 100th anniversary in 2011, set up a special domain just for the site. It's a fraternity that is very popular in the United States among black men in colleges. It supports many outstanding black men among its members, including Congressman Clyburn, Bill Cosby, Shaquille O'Neal.

And we became interested in what kind of advertisements showed up on that site. Well, there are lots of ads about graduate degree programs, which, of course, seems incredibly appropriate, given that this is an undergraduate fraternity -- and a clicker that doesn't work. What is it with this?

(Laughter.)

MS. SWEENEY: There are also advertisements about, you know, luxury vacations and other kinds of opportunities like that. And then there are also these kinds of advertisements, such as this one, "Click here to view your arrest record now."

Now, there has been much said about Instant Checkmate and this is an Instant Checkmate ad. I did earlier work about the suggestive nature of arrest record ads around Instant Checkmate, but I think it's very clear to see that this actual ad is not showing up the way it regularly showed up. It actually shows up with flashing colors, so it has kind of a neon effect. But flashing your arrest record would be a presumption that this particular audience would not appreciate.

It wasn't the only ad, though, that made that kind of presumption. There were also ads for a criminal lawyer and there were ads for credit cards.

Now, it turns out the financial industry is the number one marketer online. So they're the number one industry that's advertising online. And given what we had just seen of Omega Psi Phi, we became very interested in what kind of credit card ad is that and what are credit card ad experiences. I hope you have better luck with the clicker.

MR. ZANG: All right. So, going more generally from the Omega Psi Phi anecdote, we first started looking for word lists of quality cards versus ones that are more harshly criticized online.

So, here you can see a list of the top 25 most harshly criticized cards or the most highly praised cards that we were able to find. And for Omega Psi Phi, they actually had two of the ads from the harshly criticized list show up on their site, including First Premier Card and the Centennial Card. None of the ads from the highly praised cards list actually showed up on their site. And, in fact, for the highly praised cards list, it's not necessarily those cards are all just high credit score, really luxury cards. In fact, you had secure cards that were highly praised as well, like the Capital One secure card.

So, but digging back into the comparing of the two cards, what we saw was if you looked at the most popular ad that ran for a First Premier card, which is one of the most often criticized cards, if you go online and compare that to the most popular ad that was run by American Express for their blue card, the sites that appear that those card ads appeared on do look very different.

And one theme that quickly jumps out at you, especially for the American Express blue card is, around higher education, where you had sites such as Harvardmagazine.com or Yalealumnimagazine.com or, like, the Heismanwinners.com as sites that American Express is advertising on.

On the other hand, for First Premier's card there didn't seem to be as much of a cohesive theme that we picked up.

MS. SWEENEY: So we wanted to dig further. Like, what is the nature of these cards, where are they appearing generally, and is it somehow related perhaps to the popularity of the website.

So if you think about popularity of websites, there are a few websites that are highly popular, almost everyone goes to, they're on the top of everyone's top ten list. And then the popularity of the website drops as you go further out.

Alexa is a company that ranks the traffic to and from domains and so we used them to rank all of the publishers of all of the credit card ads' deliveries that were made of the praised cards and the criticized cards. And what we learned was that the criticized cards appeared completely across the entire spectrum in increasing order as the popularity of the domain drops. So it's a curve that's going this way. And in every segment of the popularity zones, there are, in fact, credit card ads for the criticized cards.

The highest number, though, were in those ads whose popularity ranks were above a billion. Now, to be above a billion, you probably aren't getting much traffic, that would be the issue with respect to their popularity. Those ads that are close to the left are highly popular, those are very curated, and there's a lot of information that exists about the audience. And you could actually look up on services with Quantcast to find out the demographic make-up of those websites. But when you're way out in the billions and millions, that kind of information doesn't exist.

The other thing to note, though, is where were the praised cards? They didn't follow the same pattern. Instead, they were heavily generated in the middle around the 100,000 to the one billion.

So these ads are showing up on different popularity, on websites and domains with different kinds of popularity profiles.

MR. ZANG: And another perspective that we took to look at the type of sites that these card ads were running on was from the perspective of understanding that different websites do attract different types of audiences and that there are websites out there that are more exclusive to an audience of one demographic group than other demographic groups.

So we took the approach of analyzing Comscore's data on the browsing behavior of 46,000 American households in 2013 and looked through the four million websites those households go to, to look for sites that are more commonly visited by households of certain demographic groups.

And so, for example, if we took a racial lens to demographics, you can -- we found that for Latino Americans they're more likely to go to sites like Univision or Tarango or Musica.com. For African American households they went to sites like Worldstarhiphop or Footlocker.com.

Now, in this case, it doesn't necessarily mean that only African Americans go to Footlocker.com.

Footlocker.com could have lots of other visitors from other racial groups as well, but African American households are much more likely to go to Footlocker.com.

And so we looked at exclusivity from the lens of race, from age, from income, from the level of education in the household and also whether the household had children or not. And we are able to find for each of those different lenses, sites that were exclusive to each of those groups.

And this raises a question for us of if there are sites that are out there that are more exclusive to certain groups, what is the advertising experience like on those sites. And could there be the potential for disparate impact if -- depending on the type of ads that are shown or the type of ads that are actually not shown on those sites.

MS. SWEENEY: So, one of the things that we learned was that these groups are appearing almost evenly across the entire popularity of these domains. That means that no matter which ad campaign you ran, whether one you were trying to focus on popular domains or less popular domains, you could easily encounter one of these domains for which there was an exclusive audience, because, in fact, they appeared in all the domains.

So, what we then became interested in was to what extent could we predict whether or not the ad would receive or were there sites in the comScore data that should have received these credit card ads and were those sites part of these exclusive groups.

And we found that it’s true, that around race and income and age, there were differences. And, in fact, there were praised ads. And these praised ads, for example, for Asians, we saw Capital One secure card and Capital One and CitiBank and Discover finding domains for which Asians -- were more exclusive to Asians. And you couldn’t tell by the name or the key word of the page, the domains are names like Dealstobuy.com or Visajourney.com.

Discover did a very good job using Seekingalpha.com to target people who are more -- whose income is $100K or more. That’s an exclusive audience at Seekingalpha.com and it’s a very popular site.

And then we also found examples in age ranges. Discover, with ages 18 to 20 and Capital One secure card found some domains that were somewhat exclusive to ages 25 to 29 or ages 65 plus.

So, domains with exclusive audiences do exist and ads are not exempt from being delivered to those sites. So the lack of ads or too much of another ad, could lead to a disparate impact. And demographics could, therefore, sort of infer what kind of advertising experience might you have.

We're going to stop here. If you want more information about the work, we'll have a blog post later with some of the details and a paper to follow right after that.

I did want to leave the panel that's coming up next with three questions from this work. One of them is that by subscribing to an online ad network a publisher may not have an opportunity to review ads anymore and if there is a problem, what are the publisher's rights and responsibilities?

Another question that comes from this is when we look at Omega Psi Phi. What are the sufficient and necessary circumstances for a community to experience adverse impact in this setting?

And the last question is that the kind of audience exclusivity measure that we used to find these audiences that had that type of exclusive nature to the audience is something that could actually be used inside of the big data engine in that same fraction of a second, to realize that this ad probably shouldn't go to this site at this time.

If that's so, and it's that easy to do, should or how might a big data analytic engine be required to use it or an equivalent remedy?

So to find out more about the work, check out our Tech@FTC blog. Thank you.

(Applause.)

MS. WORTHMAN: Good afternoon. My name is Katie Worthman, I'm an attorney in the Division of Financial Practices here at the FTC and I am co-moderating the third panel, along with my colleague, Patrick Eagan-Van Meter, who is a program specialist, also in the Division of Financial Practices.

Panel 3 is titled Surveying the Legal Landscape. And today we are going to look at the various anti-discrimination and consumer protection laws that impact big data.

Let me first quickly introduce the panel. To my immediate left is Leonard Chanin, who is currently a partner in the law firm of Morrison Forester, who in a previous life also was head of regulations at the Federal Reserve and at the Consumer Financial Protection Bureau.

Then there is Carol Miaskoff, who is in the Office of Legal Counsel at the Equal Employment Opportunity Commission.

Montserrat Miller, who is a partner in the Privacy and Consumer Regulatory, Immigration and Government Affairs Practice Groups, at Arnall Golden

Gregory.

And Lee Peeler, who is President and CEO of the Advertising Self-Regulatory Council and Executive Vice President National Advertising Self-Regulation Council of Better Business Bureaus.

And then last, but definitely not least, is Peter Swire, who is a Professor of Law and Ethics at the Georgia Institute of Technology, as well as Senior Fellow at the Future of Privacy Forum and the Center for American Progress.

And with that, I will ask Patrick to open up the panel with the first question.

PANEL 3: SURVEYING THE LEGAL LANDSCAPE

MR. EAGAN-VAN METER: So, Panel 2 kind of teased us a little bit with the laws that might apply to the big data space, so I wanted to ask all of you what you think the Federal laws that touch on the collection and use of big data are.

MS. WORTHMAN: Leonard?

MR. CHANIN: So I was asked to give a little background on the Equal Credit Opportunity Act and Regulation B, just to kind of do some level setting in terms of how that law may apply to big data marketing and those sort of things. So I'll spend just a couple of minutes talking about that.

So the Equal Credit Opportunity Act implemented by Regulation B, the Federal Reserve Board administered that regulation for many years and it was recently, or a couple years ago, transferred to the CFPB. So, interestingly enough, the Equal Credit Opportunity Act doesn't apply to marketing activities, Regulation B does to a limited extent. And the reason is the law says it's illegal to discriminate against an applicant in connection with a credit transaction. An applicant is defined as someone who has applied for credit. So if you have not applied for credit, technically speaking, the law does not apply to you -- that is, the Equal Credit Opportunity Act does not apply to pre-application activities.

The Federal Reserve Board, though, many years ago applied Regulation B to certain activities at the pre-application stage, but it's pretty narrow or focused, if you will. So, first of all, the law says you cannot discourage a person from applying for credit on a prohibited basis. And that means something like you cannot make statements to a person, you can't use advertisements, radio, newspapers, and so forth that would put forth symbols or tags that would discourage a reasonable person from applying for credit.

The second way that Regulation B might cover marketing activities is if you're an existing account holder. So there you have a credit transaction with the lender and the lender cannot make statements that would discourage you from using your credit or provide different terms to you, since you are, indeed, someone part of a credit transaction.

So, generally speaking, Regulation B applies to transactions or applies to marketing in those relatively focused ways. But it's not a new issue that we're talking about in terms of marketing. In fact, the Federal Reserve in 1985 looked at marketing activities, decided at that time not to expand the regulation scope to cover marketing activities. And again looked in 1998, when it was reviewing Regulation B and solicited comments on whether pre-screening activities should be covered by Regulation B.

In 1999, the Federal Reserve Board decided that it was not appropriate to apply Regulation B in its full context to pre-application activities, marketing activities, because it did not have evidence that suggested that lenders were using, in any significant way, prohibited bases for marketing. The Fed also said, though, it had anecdotal evidence suggesting that some lenders were using age, that some were using geographical information in terms of marketing activities, but balanced that anecdotal information against the benefits of marketing. That is, that pre-screening, in particular, makes credit available to individuals. There was evidence that the Fed cited that said that allowing lenders to engage in pre-screening without coverage by Reg. B could make credit available to more individuals.

The Federal Reserve also noted that, of course, lenders could use information to discourage people from a fine, could use information to provide products to some areas and not to, if you will, disadvantage products.

So, in 2003 the Federal Reserve Board actually adopted a rule dealing with pre-screening marketing activities coming out of a 1999 proposal. That rule is still in place today. It basically requires creditors to retain information about pre-screening activities, that is, activities where creditor use is governed by the Fair Credit Reporting Act, uses credit report information, and a creditor must retain the information used to market -- that is, the criteria.

So, today, and since about 2004, if a creditor uses information -- race, ethnicity, age, gender, et cetera -- to engage in pre-screening it must, under the law, retain that information. The Fed, at the time of adopting this rule, said that enforcement agencies could use this information to determine whether or not pre-screening was being engaged in by lenders in an inappropriate fashion. Whether that data has been provided by the Fed or the CFPB in the last few years is questionable, but there is some law in place now that will at least arguably provide some more data to various agencies, in terms of pre-screening and marketing activities.

So that's a very long answer to your short question.

MS. WORTHMAN: But one thing, also, is that the Equal Credit Opportunity Act applies in what space?

MR. CHANIN: So the Equal Opportunity Act applies only to credit transactions, but it applies quite broadly, that is, to all credit transactions; consumer credit, business credit, credit to corporations, to sole proprietorships, partnerships and the like. It's not limited to consumer transactions. It is quite broad in its applicability.

There are obviously, in addition to the anti-discrimination provisions, rules dealing with adverse action notices if you decline credit to people, you have to give them a notice, and those sort of things. So it's quite broad in terms of its scope.

MS. WORTHMAN: And, Carol, in the Title 7 context?

MS. MIASKOFF: Right, right. In the context of the Civil Rights Act of 1964, as well as the other Federal EEO laws, the Americans With Disabilities Act, the Age Discrimination and Employment Act, and the Genetic Information Non-Discrimination Act, not to miss that, we have really very settled law. I mean, it's the 50th anniversary of the Civil Rights Act this year, but settled law with some basic principles that, I think, can definitely be translated into the big data space.

Now, how do these employment and non-discrimination laws sort of reach over? How does employment meld with the big data? I think it does in the spaces of recruitment, clearly, for the kind of advertising issues we've been seeing discussed here. And in areas of screening people for jobs once they have been recruited and making that ultimate selection decision. There's a real potential here, I think, to gather information about successful employees and then turn around and use that to screen people for employment.

With the screening piece, I think the issue really is about what prejudices are built into the data and, therefore, would be built into any rules deduced from the data. And, therefore, be used to select people who meet those same rules. So, would it exacerbate, perpetuate, past discrimination? I think that's the big concern.

In recruitment it's the same issue around advertising that we've seen in the commercial space. And I know -- you know, think about LinkedIn and all the jobs that may be referred to you there. And, you know, you always wonder, you know, who's getting which jobs, are they equitably distributed or not, or are they targeted?

So, I think that's the big picture. In terms of the law I just want to make a few quick points.

It's been interesting to me, because everyone's been talking about disparate impact and adverse impact a lot. In the employment space, those are very precise legal terms. And there is a cause of action for disparate impact and I would say that that's the one, frankly, that's most suited to big data, because what that's about is taking a neutral, i.e., like, race neutral, gender neutral, et cetera, term, that nonetheless disproportionately excludes members of the protected group. And -- and this is the critical piece here -- and is not job-related consistent with business necessity.

Now, in terms of big data I think this is the rub, this is really what's very fascinating, is that the first step is to show, is to look at what is the tool. Now, you know, this could apply to recruitment or to selection, perhaps more to selection.

What is the tool, does it cause a disparate impact, and once you get there, is -- you know, just because it causes disparate impact, doesn't make it illegal discrimination under the employment laws. It's only illegal if it does not predict, accurately predict, success in the job. Okay?

So this raises all kinds of fascinating issues with big data analytics, because, indeed, if you do possibly have prejudices and prejudice is built into the data, something might be validated as predicting success in the job, but it might just be predicting that, you know, white guys who went to Yale do well in this job.

So, you know, there's going to be a lot of interesting, I think, thought that needs to be done and technology work, really, around understanding how to validate these kind of concerns.

MS. WORTHMAN: Montserrat, with respect to the Fair Credit Reporting Act?

MS. MILLER: Sure. So, I am going to talk a little bit about the Fair Credit Reporting Act and try to weave that into big data and how certain reports are used in that context.

So, FCRA, enacted in the early 1970s, a consumer friendly statute. And what it seeks to regulate or who it seeks to regulate are consumer reporting agencies, so credit bureaus or background screening companies. And it’s very specific in what it seeks to regulate and how it seeks to regulate it. So that consumer reporting agencies operate in an environment in which they -- with respect to confidentiality, accuracy and then also the legitimate use and permissible use of data.

And when you’re talking about the FCRA or the Fair Credit Reporting Act, we’re looking at consumer reports, consumer reporting agencies, users of the consumer reports and also furnishers of the data for the consumer reporting agencies. So it’s an ecosystem in which these companies operate under the Fair Credit Reporting Act.

With respect to the reports, themselves -- and this is where you begin to get into, obviously, the data, the reports could include credit, they could include criminal history information, obviously, that’s something that comes up with employment, in both of those they could include drug testing information, employment education verification, public records information.

So these are reports that are put together by consumer reporting agencies provided to, for instance, employers, landlords, and others, all within the confines of the Fair Credit Reporting Act. And the information contained within those reports, the data contained within those reports, goes to credit-worthiness, standing, capacity, and it also goes to character, general reputation or mode of living.

And Commissioner Brill covered some of these points already, so I won't belabor them, but as she mentioned, they're certainly looking at the use of that data for credit or insurance or employment purposes, or other purposes. But all purposes which are defined and regulated under the Fair Credit Reporting Act are permissible purposes.

So, I would say with respect to the Fair Credit Reporting Act, you're looking at, as I said, permissible purposes, due diligence of end users, who are going to be looking at the data, consumer reporting agencies must operate with maximum possible accuracy, and there's always, and most important for consumers, whether it's for employment or tenancy or credit or insurance, other purposes, there's always the right to essentially appeal and challenge the accuracy and completeness of any consumer report.

FCRA, over the years, has not operated on its own. We've certainly seen the states coming into this space and especially, I think, aggressively over the last few years when you talk about the potential discriminatory impact of the data that's in those reports, and, really, with respect to credit and criminal history. So, you have not only the FCRA, which is enforced by the FTC and the CFPB, and also there are private rights of action, but you have the state analogs, which are essentially their own mini FCRAs and you have California, Colorado, Maine, Minnesota, New Mexico, New York, Oklahoma, Vermont and Washington State.

So you can see there are a lot of people, a lot of different entities, Government entities, enforcers, that are operating in the space of using this big data with respect to the permissible purposes.

And then you also have other states, which have gone more -- in a more limited, but important, area and consider whether the use of credit for, say, employment or tenancy might have -- that in certain settings the use of credit could have -- be considered an unlawful or discriminatory practice. And the same applies with criminal history information.

So, certainly, there are examples of states who are very active in this space of data, big data, and how it's used in these reports, in seeking to protect the community and I think certainly some individuals in certain communities.

MS. WORTHMAN: Thank you. And, Lee, could you speak a little bit about section five?

MR. PEELER: Sure, I would love to. And, although, I do think I was sort of targeted as the legal historian on this panel.

And, you know, I do want to also just commend the FTC's leadership on this. You know, data is now the economic lubricant of a lot of our economy. And looking at this issue is in the finest tradition of the Federal Trade Commission -- in fact, if you're looking for historical analogies, in the 1960s the FTC launched a ground-breaking review of inner city retailers' marketing practices and that led, under the Federal Trade Commission Act, which I'll talk about in two seconds, to a whole wave of initiatives that really changed a lot of what we traditionally thought of about credit practices and debt collection practices and merchandising.

So, you know, I think this is really, again, in the greatest tradition of the Commission.

I do want to go back to some remarks that were made this morning, though, and say I think you have to -- you can't just look at the application of the FTC Act broadly on big data. I think the remarks that were made this morning really say you have to look at how -- at where big data is being used and how it's being applied.

And so one critical distinction that I think has been talked about a little, but I think is really important for what I'm about to talk about, is the distinction between decision making, granting or denying credit, granting or denying a job, and advertising and marketing. And, you know, the decision making for credit, longstanding prohibitions, going back to 1974, on using, you know, marital status or race in the decision making.

In advertising the traditions are the opposite. Advertising is necessarily about targeting your products to markets. You can just look at cosmetic ads, if you look at ads for shavers, if you look at ads for music, for books, all of those ads you're going to find targeted. And probably, you know, the best example of ongoing massive targeting is in selling political candidates right now.

So how does the FTC Act apply to those areas. And, again, my background is advertising, so I want to focus on advertising in talking about the application of the FTC Act.

The first piece of the FTC Act is deception. Whether an act or practice would mislead a consumer acting reasonably under the circumstances. And there were sort of two basic applications there. One is well established legal principle, if you're narrowly targeting an audience, you're responsible for the reasonable interpretation that audience would have. So if you're, you know, targeting your ads to cancer patients in a well-known FTC case, you are liable for what the interpretation of that ad would be and what information that consumer would need, if you're narrowly targeting.

The other example that I think will be important as the FTC goes down the road is, you know, data brokers are responsible for the accuracy of what they tell consumers and tell marketers they're providing them, so they're responsible for the accuracy of the representations they make about their database.

The second core aspect of FTC jurisdiction is unfairness. There is a long history of unfairness that led to its codification in 1994, but it's a -- you know, it's a provision that's been in the Federal Trade Commission Act since consumer protection authority was created in 1934.

The elements of an unfair practice under the 1994 codification are that the practice is likely to cause substantial consumer injury. And that that injury is not reasonably avoidable by consumers and on that particular part of the analysis you would need to look at whether the ad is targeted to a specific group, but also what's the consumer group's access to alternative products, how easily can the group go on and find alternative products at better prices or at better terms?

Even if you met that analysis for advertising, your next -- the next challenge is to show that that harm, that net harm, is not outweighed by benefits to consumers or competition.

And again, you know, a flat ban on use of, for example, gender in advertising would probably fail under that approach because, you know, take, for example, an entrepreneur wants to open a women's shoe store. They will be targeting their ads based on sex and gender.

Probably -- and then the big issue for legal analysis under section five is what extent has well established public policy had. And we have a very well established public policy in the United States of not treating people differently. The statute that created the codification is quite clear, that you can use public policy in weighing the costs and benefits, but it cannot be the primary basis for the conclusion that the practice causes net consumer injury.

And then a last -- two last pieces of the FTC's authority that I think are really important for the discussion today, is what you're doing right now, which is the ability to use your 6(b) authority to collect information, issue reports and inform the public about what's really going on in the marketplace is invaluable.

And the last is not a specific provision of the Federal Trade Commission Act, but the FTC's programs of educating consumers. And as Commissioner Brill said earlier today, really encouraging industry to step forward and educate consumers themselves.

And then the very last point I want to make is I thought Commissioner Brill and Montserrat did a great job summarizing the Fair Credit Reporting Act. But because I'm an industry self-regulator I -- when I first got to the FTC I took the Fair Credit Reporting Act as it existed then and you could almost -- it was almost verbatim from a pretty well established set of industry self-regulatory principles that had pre-existed the Act by several years.

And the only lesson I -- the important lesson, I think, to learn from that is that by looking at what the industry is doing on a self-regulatory basis, you can come up with workable -- you're more likely to come up with workable solutions to issues, than if you just try to create it yourself.

So that's my summary.

MS. WORTHMAN: I'd like to turn a little bit, Peter, to an example that was used this morning on Panel 1. And it was the Maserati example, where apparently

Maserati, the sports car, the example that was used by Mallory Duncan was that the dealership has information that the Maserati is most likely to be sold to this list of people. There's a 30 percent chance that people who get any type of offer will come in and purchase the Maserati. And the list happens to be 95 percent male.

So, the question is, does that -- if you send a flyer advertising a free test drive to this list that's 95 percent male, does that implicate the ECOA, fair marketing purposes? Does it matter if it's Maserati Finance Company?

And I know that you've released a paper recently on fair marketing.

MR. SWIRE: Okay. Thanks. So I'll briefly say that and then make a couple of other points that were in the paper perhaps.

One of the things in the disparate impact test, which is the way the Equal Credit Opportunity Act has been applied, is it's facially neutral, but then if there is a different impact on the protected class, is there a business necessity and is there any less restrictive way to do it.

And you can certainly imagine where the act applies. That advertising to women's shoes or for the Maserati, if the facts are there, there'd be an argument that there's business necessity and then there'd be a question of is there less restrictive alternatives.

So that's the way I think it's been done in the fair lending context. I would like to, just from the paper, make a couple of points because I think we've heard some reasons for caution in thinking that there's claims here from the plaintiff's side. And there's also some reasons to think existing law has some teeth that haven't been brought out.

And so the first one, I think, is -- and in interviewing people who do fair lending compliance, there are huge fair lending compliance programs. The level of effort in the major financial institutions in this area is very large. And at least part of the reason is related to a CFPB case in June this year, where GE Capital was ordered to provide, or did a consent decree to provide, $169 million in remedies for fair lending violations in advertising. And that's just a big n umber compared to what we're used to in consent decrees and such.

And the facts were about advertising to existing customers. As Leonard pointed out, it's especially clear the law applies to existing customers, but according to the facts in the complaint, GE Capital had offered a nice credit deal, you can reduce your back amount that you owe, but it did not extend those offers to any customer who indicated they preferred to communicate in Spanish or at a mailing address in Puerto Rico.

And so the violation was that you only advertised in English, you did not advertise in Spanish, you were excluding Spanish-speaking consumers from this very attractive offer and as a result, you know, $169 million consent decree.

And I think when you talk to fair lending people, they're aware of ways the law may or may not apply, but they're aware of that level of enforcement and it gives them a different level of seriousness.

And so from seeing cases like that over the last 20 years -- that was an unusual one, but cases that have been brought in, I have three very quick points. The first is the FTC has unusual enforcement power under the Equal Credit opportunity Act, so the statute specifically says the FTC can enforce compliance with it, irrespective of whether that person is engaged in commerce or meets any other jurisdictional test under the FTC Act.

So, for those of you who have been afficiandos of the FTC enforcement jurisdiction, this is a sort of spectacularly interesting moment in the law that I think is worth noticing. It doesn't have to be somebody engaged in commerce and so there are some important FTC powers here that are not familiar from other statutes.

The next one is -- as we wrote this paper and tried to think about fair lending and its history, which is something I worked in a while back, and how it makes sense to privacy people, many of whom are in the audience.

The first point is that there is sectoral legislation in anti-discrimination law. And that's really familiar to the HIPAA, Gramm-Leach-Bliley, COPPA sectoral regulation in privacy. And so we have the ECOA, the Fair Housing Act and you have Title VII, so there's existing substantial legal laws in place around lending and housing and employment.

And so, for those areas, it's sort of like HIPAA and Gramm-Leach-Bliley, it's time to go do the research and see what those laws cover or don't.

And then the last point I'd make is -- similar again for privacy people, those are the HIPAA, Gramm-Leach-Bliley, COPPA regulated parts. And then what do those principles teach us about everyone else? And I think in privacy those laws have been looked at as the structures that people use for a lot of their privacy policies in other areas. They may or may not have all the strictness, but it's the same structures.

And so I think the last 20 or 40 years of discrimination law, including fair lending, provides a lot of useful insights about advertising and other practices related to big data. And instead of these issues being brand new -- and this is something that Leonard said -- they've been going back to the '80s and '90s, we have decades of work that's been done here. And I think along with figuring out what we think we ought to do, there's a legal research task about what the law has done. And talking among others, fair lending and employment and fair housing experts to see what's really done there is something that I think really would inform our debate a lot about what the legal rules are.

MS. WORTHMAN: Now, going just a little bit into the -- not to beat the Maserati example, but let's say that the list is based on aggregate data that has been prepared by the credit bureaus, on a household level, not on an individual level. What are the implications in the FCRA context for a marketing list that has been prepared using previous purchasing history by consumers just by households? Does the FCRA apply in that context?

MS. MILLER: Well, I will -- I think I'll punt on this one, because I think marketing is not necessarily my expertise. It's more FCRA and consumer reporting with respect to other permissible purposes.

MS. WORTHMAN: Anyone else would like to take it?

MR. PEELER: So just as a general principle -- and I actually had the opportunity to work on implementation of the Fair Credit Reporting Act and to work on Reg B when it was issued. And I think just looking back at the structure that's there, if you were using information collected from a third party to make decisions about whether an individual can purchase or obtain particular good or services, I think you do need some structure to provide FCRA type noticing correction, as opposed to if the issue is are you sending an ad out. And I think, you know, one of the things I think is true, Leonard, still is that pre-screening, where you have exercised jurisdiction, still involves making a firm offer of credit, right?

MR. CHANIN: That's correct.

MR. PEELER: So, again, just looking at the model that's been used for years and years in that industry, if you're making a decision about, you know, what's going to exclude somebody based on third-party information, there ought to be some way to make sure that information is right.

If what you're doing is just trying to make information available to consumers, I think that the cost of doing a fair lending type analysis for, you know, a wide variety of products gets to be, you know, very high and very unworkable. And the exception to that, I think, are two of the areas that are represented up here, housing -- you know, that's a limited commodity. If you miss the opportunity to apply for housing, you know, you're not going to get the housing. Jobs is a limited commodity, if you miss the opportunity to get your application in for that job, you're out of luck, you can't come back and get, you know, the extra one of those.

MR. EAGAN-VAN METER: So based on the research that LaTanya presented today, would marketing high interest rate/low credit limit credit products on certain websites, based on the consumers who frequent those sites, implicate any of the statutes we discussed today or any others?

MR. CHANIN: So I'll take a first jab at that. If you're not talking about housing, but you're talking about other credit, then I think generally speaking the answer is no. First of all, at least my assumption is that someone, anyone, can apply for credit. That is, that if I market it, it’s not the sole way or the only way to get credit, because that would raise other issues fundamentally, whether you're discriminating if someone can't call you, go on your website, however you can apply.

But assuming that you market and people can contact you independently of that marketing activity, then I don't think that marketing and target marketing would be -- would raise fair lending issues, at least under the Equal Credit Opportunity Act.

I think -- you know, as was alluded to by Lee and others, you know, this is not a new issue. That is, people for years have been targeting marketing in radio, television, newspaper subscriptions and so forth, in order to get people who might be interested in their products, whether credit products or other products to respond to those. What we've got now is obviously far more data that people are able to use and manipulate it in order to better target, if you will, to audiences that they think may be interested in their products.

The other thing I'll mention is that, you know, it's been -- I wasn't able to attend this morning, but there's a lot of discussion about disparate impact. If you decide to apply the Equal Credit Opportunity Act to credit, you need to talk about disparate treatment. What the law would prohibit is if I have, for example, as was mentioned earlier, a retail shoe store predominantly or exclusively for women, in terms of women's shoes, and I offer a credit product, it would be illegal for me to target -- that is, to send solicitations to advertise solely to women, regardless of disparate impact. That is, de facto discrimination against men would be illegal if you apply the Equal Credit Opportunity Act to marketing, unless you have some kind of carve-outs or something.

MR. PEELER: And the two quick clarifications on that is if a man comes in and applies for that credit card, he's got to be evaluated on the same criteria as everybody else does.

And your credit portfolio in the credit card area is going to be evaluated against whether there's disparate impact. So the end results are important and, you know, if you're a creditor I'm assuming that you're making sure that your marketing is going to get you to the place where you can survive an examination by Leonard.

MR. CHANIN: Not anymore.

MR. SWIRE: So, Leonard has lived these issues at the CFPB in recent years and I'm in the midst of getting up to speed again on some of this, so I -- but I would like to point out two things about marketing in the lending area.

One is that in the fair lending area there is a history of strongly encouraging targeted marketing to minority communities. So if you go and look at the remedies, the answer is you haven't been advertising on African American radio stations or you haven't been advertising to Hispanic radio stations and you need to do that. So instead of marketing being this sort of bad thing, as you sometimes hear in the privacy debates, it's been a required part of the remedy for fair lending.

But along with that, there's a sort of split, which the paper would call the paradox of advertising on lending. And which is that there's a prohibition on what's called steering when you lend and this has been in the rules for a long time.

And at least in recent years, after the CFPB sort of saw the subprime crisis and whatever, targeted subprime loans and targeted other loans, I think, has raised CFPB concerns so here's a quote from its guidance, "A creditor may not advertise its credit services and practices in ways that would tend to encourage some types of borrowers and discourage others on a prohibited basis." This is the CFPB language. "In addition, a creditor may not use pre-screening tactics likely to discourage potential applicants on a prohibited basis."

So there's at least language that's sort of more -- if you want to call it plaintiff friendly or enforcement friendly, than some sort of categorical idea that this is exempt from the ECOA. And it may be the CFPB is pushing past some of the previous ways that people thought about it at the Fed in earlier years. But there's language that's more pro plaintiff than some of the categorical exclusions would suggest.

MR. EAGAN-VAN METER: So to push that a little bit further, if you had a high end credit card and a more sub-prime card, and the sub-prime card was only marketed on sites frequented by minority groups, and the prime card was on, you know, sites that were frequented by high income or, you know, other nonprotected classes, does that count as steering in that way? If you're not, you know, kind of turning someone off, but you're giving them a different offer that might not be as appealing?

MR. SWIRE: Is that --

MR. CHANIN: You raised the steering issue.

MR. SWIRE: I did. And in the pre-call with Leonard, I said, Leonard, even at the CFPB when is it good to do a targeted marketing to make up for past problems and when is it bad to steer and can you point me to the authoritative source on that? And we weren't able to identify an authoritative source.

So I think this is a real puzzle. And my paper suggests it needs a lot more discussion than we've had today. But maybe, Leonard, you have more?

MR. CHANIN: Yeah. I guess what I would say is the fact of marketing those products to different either audiences or different websites, in my view, does not violate the Equal Credit Opportunity Act.

However, as I think Lee alluded to earlier, if your portfolio -- if you have data and, you know, sometimes lenders do not have this data, but if you have data that shows ethnicity or gender or age, and so forth, in those portfolios then certainly there are going to be questions about why do you have such a skew in terms of who has these credit products. Do you make them available to everyone? If someone calls up, goes on your website and applies, do you steer them? That's going to raise very different issues.

But the fact that people respond to certain ads and other people respond to different ads, I don't think raises that type of issue. It's simply, what does the portfolio look like at the end of the day and how will you explain those, if there are dramatic differences.

MR. SWIRE: Can I follow-up just on -- so, it was interesting what Leonard said, if you have the data in your portfolio that indicates a skew, that's reminiscent of having HMDA data, Home Mortgage Disclosure Act data, that shows a potential skew and then regulators historically have looked more carefully at it.

The paper I wrote suggests that that data about likely demographics may well be available in online marketing in a lot of ways it wasn't historically for lending. So a lot of online marketers are pretty sure they have a pretty good fix on their market and so there may be data inside their big data sets that say with some level of confidence what are the demographic, you know, characteristics.

And if you have that and you have a disparate impact in the data in your database, the history under fair lending has been that you might come under scrutiny, at least for the regulated industries, I think.

MR. PEELER: Well, and I think one other risk would be -- if in the hypothetical you raised, if somehow when -- if the consumer goes back to that creditor, not in response to the ad, but goes back to the creditor site and somehow the products that that consumer is able to access on the website is limited to products that fit a particular profile, then I think you probably do start to engage -- have some serious issues.

MR. SWIRE: Let’s call those landing pages that might be different for customers of different sorts.

MS. WORTHMAN: Going to -- actually, following up a little bit on some of the panel discussions from 1 and 2, they discussed aggregate credit scores. How is the industry applying the FCRA analysis to these scores?

MS. MILLER: So I'll say -- I'll start that one. With respect to employment and the FCRA and the use of that data for employment screening purposes, there is a common misperception that these credit scores are used for screening purposes and they're not.

And so, therefore, if you were to request a report on an individual and you're a consumer reporting agency and you're providing that to an employer, it's not going to include a credit score. It may include credit information, but it's not going to include a score.

So, taking that off the table, although I know that there's a lot of -- the media certainly reports at times that scores are used for employment screening purposes, in fact, they're not used for that purpose.

Now, I think that if you have just the general aggregate scoring and you're looking at certain communities, I think then it would turn more to a discussion about the discriminatory impact of the use of that type of data.

MS. WORTHMAN: So how is that implicated, Carol, with the fact that when somebody applies for a job they can definitely give their consent to have the employer look at their credit history. But even if they're following the FCRA, how does it impact with Title VII?

MS. MIASKOFF: Right. Well, even if someone gives their consent to doing, you know, getting the credit background, if the employer uses it as a reason for excluding someone from employment and if using that has a disparate impact, and is not -- and the key is, is not job-related and consistent with business necessity or even if it is, there could be a less discriminatory alternative. In that case, it's going to be discriminatory, regardless of the consent. So that's the bottom line there.

MS. MILLER: And I just wanted to piggyback off of that. I mean, certainly, consent is the first step in terms of pulling such a report for employment screening purposes. I will say also that credit is not as frequently used as one would believe that it is. There are other -- there's other data in the reports that is more frequently used. And credit tends to be very specific to a position, which would blend nicely with Title VII and what Carol was talking about.

But that's the baseline, is you have to have the individual's consent.

MS. MIASKOFF: I would just add that as a very practical matter there are probably not many employers out there, looking at the whole landscape, who understand how to read the kind of information they get when they get one of these financial reports about someone.

And I think probably that's why everyone talks about credit scores, because that's something that a lot of us can understand. But when they get a lot of other information it's often hard for them to put it in context and, you know, therefore, an employer might just say, oh, we got a hit, you know, we have something. And then potentially exclude someone.

MS. MILLER: Which I would say is why credit, with respect to employment screening, is used sparingly and scores are not used. In fact, there are contractual restrictions to the use of scores if your permissible purpose is for employment screening when working with one of the bureaus. And certainly with the reports themselves, it is important to understand what they say. And there are also, even at the state level, quite a few restrictions on the use of credit if it is, in fact, for employment screening purposes.

So I think credit is an area that is highly regulated, whether it's FCRA or state statutes.

MS. WORTHMAN: And then taking a question from the audience, Carol, you said earlier that big data, if it has a disparate impact but it's predictive of job-related outcomes, that it's not illegal. Does that mean that the better the data set the more likely it is to comply with the law?

MS. MIASKOFF: I guess the more likely, yes. But whether or not it, in fact, complies is the actual question. And the issue is going to be, just to sort of clarify, really whether or not the criteria used to screen someone out for a particular job, you know, is relevant for performing that particular job.

And I guess I didn't mentioned before, but one of the ways in which you could say the EEO laws anticipated big data, is that we have at this point, quite -- from 1978 some guidelines in place about validating selection tools for employment. And they were written initially about tests. And the question was if a test had a disparate impact, how do you know if it's job-related for the position in question and the tasks in question.

And it has three ways of validating. And I think it's really going to be interesting to see how those principles can be applied in the big data context. But it is for the job in question.

MR. SWIRE: Can I follow-up on that?

MS. WORTHMAN: Yes.

MR. SWIRE: So there's a Sears case with employment about -- it turned out men were more likely to do certain high commission sales and women were more likely to be near the front of the store selling smaller items.

And Sears was able to come up with a statistical study in the case that showed a business necessity that it was actually based on the choices of the individuals who had picked these different jobs. In that case Sears won, the defendant won, but it won after having a pretty substantial burden of proof to show the validation on the statistics.

MS. MIASKOFF: Yeah, it's not easy.

MR. SWIRE: And so I think in the marketing area, the fair marketing or whatever we call it, one of the changes, if this law turns out to apply in these sectors, may be that the practices meet business necessity, but there would be a compliance effort by the companies to show the validation. And I think up until now that effort to do that validation has not been the industry standard in a lot of places. And to meet the laws, it might become or have to become the industry standard.

I'm curious from the employment side, does that match your understanding of the law, at least in the employment side?

MS. MIASKOFF: Well, as a -- I wouldn't say it's more whether it matches my understanding of the way businesses are complying with the law or not.

MR. SWIRE: Right.

MS. MIASKOFF: The reality out there is I think Federal contractors, because of all of the requirements that come with a Federal contract, do a lot more validation now than companies that are not contractors.

I know from EEOC's perspective, we regulate all private sector employers with 15 or more employees and one thing we really are pushing now is the kind of record-keeping that can facilitate validation.

MR. SWIRE: But there may be a due diligence effort here expected from the companies that has not maybe been built in, up until now.

MS. MIASKOFF: And I think that could be a very positive thing, actually.

MS. WORTHMAN: Now, going to another example of the use of, or the potential use of, big data. So, in 2008 the FTC brought a case against a credit card marketing company that was looking at the shopping habits of its consumers and actually based on where the consumers were shopping decided to lower the credit limits of certain consumers and actually then charged over-limit fees as a result of that.

But now, since there is this proliferation of information where you can purchase data of where people shop or use that, what are the implications, for example, if a creditor would offer credit terms, better credit terms, to people who shop at Walmart versus 7-Eleven. Or, if in the employment context, if an employer was relying on these sort of marketing lists to determine who they would advertise jobs to or who they would hire?

MS. MIASKOFF: Well, I'll just jump in starting with employment. The question would be whether -- you know, we'll look at the data and is that causing a disparate impact on one of the basis protected by the Civil Rights Act or one of the other laws. And if it did, then if it were not job related consistent with business necessity, it would be discriminatory.

MS. MILLER: And from the FCRA perspective, what I would look to in that type of a situation is just who is preparing the reports and what's being included in those reports. Because, you know, from the employment context you have to have a consumer reporting agency who is assembling and evaluating the information, providing it to a third party, and then it's for one of the seven factors and it's being used for permissible purpose.

So the question would be, do you meet all of those, do you fall within or outside of the FCRA? But certainly those are questions that come up regularly when companies try to promote new products and whether it will be, in fact, an FCRA product or not. So you'd have to look at those factors.

MR. PEELER: It sounds like your hypothetical, there actually is a decision being made about the customer, in terms of what the rate is for their credit card, so that's clearly covered by existing law.

The one sort of additional nuance that I would throw into the mix, though, is, you know, again talking about the need to segment the conversation about big data, if the information is collected online to support online behavioral advertising, the advertising industry self-regulatory guidelines say you can't use that for employment insurance or credit decisions, period. You can use it for marketing, you cannot use it for decisions.

MR. EAGAN-VAN METER: So to follow-up on that, how frequent are contractual disclaimers? Such as the prohibitions that you're referring to, kind of banning the use of that type of data for FCRA purposes.

MS. MILLER: Well, it didn't work out too well for Instant Checkmate. But, I mean, certainly it's something that -- you can't have a disclaimer, I would argue, and expect that the FTC wouldn't look at it very carefully. And especially if your disclaimer happens to be -- even though we have -- and I'll just use big data, because that's what we were talking about -- even though we happen to have big data and even though we happen to be selling it to you and even though you happen to be looking at it, and maybe, perhaps, kind of/sort of you're looking at it for employment purposes or housing, we're not a consumer reporting agency, this is not an FCRA product.

The FCRA, I think, can be -- in fact, is very effective. And I think FTC is very effective at enforcing the FCRA. So, disclaimers are certainly something that don't bode well for the company who -- especially if you're trying to say that you're not an FCRA product when, in fact, you meet all the elements of it, whether it's employment or tenant screening or if you're using, as I said, the data and you fall under the elements of what is a consumer reporting agency.

But on the other hand, you know, that is one that some would argue that the FCRA covers consumer reporting agencies, but it leaves a bit of a hole when it comes to employers who may be using that information themselves and not operating or using the services of a consumer reporting agency. So, in that situation we'd have a different analysis.

I wowed everybody into silence.

(Laughter.)

MS. WORTHMAN: What about the use, though, again going back to some of the more sort of aggregate data, nontraditional credit information, that's being used, whether it's Government records, social media, shopping habits, web tracking, location data?

If that is being used in the marketing context, both in the credit and non-credit space, is that something that is -- is there a gap there with the statues and the regulations?

MR. CHANIN: I'll take a try at it. I guess the question is -- where do I start with it. So if you think about amending the various laws, the question to me first would be is there injury, is there harm to consumers? Because you need to balance that against counter-veiling benefits.

You know, if someone is sending marketing materials based on whatever information, targeting to individuals, presumably there is some benefit to those individuals who receive it. Requiring that information to be sent to every individual, many of whom have no interest in it, is probably not going to be very beneficial, it also is going to increase ultimately the price of the product, lead to other techniques to market and so forth.

So, to me, the question is, is there injury. It seems to me there would be injury if, for example, I market through one channel or multiple channels. If the terms, as Lee, I think, alluded to -- if the terms of that credit are only available through that channel and someone contacting me through a website, through a telephone, in-person mail, cannot get those terms, they get terms that are less desirable, then that certainly could raise questions of injury.

If that's not the case, then the question to me is fundamentally are there consumers being harmed by not receiving a particular offer.

MR. PEELER: And if you expand it beyond credit, you know, you get pretty quickly to, you know, examples where it doesn't make any sense, you know, which would be the -- you know, cosmetics and shavers and, you know, music. And you also probably get very quickly to some areas where it would be unconstitutional, like, birth control or marketing political material.

MS. MILLER: And I think there's a very interesting and fascinating intersection with the use of social media. We've talked about that a lot today with -

- between the FCRA and consumer reporting agencies who are providing social media information for, say, employment purposes and then just EEO laws. Because certainly under -- employers are using social media, whether it's private employers, whether it's Government, social media is used.

And so there's sort of this split between well, what happens when an employer looks at it and Googles a candidate and then what happens when a consumer reporting agency prepares a report that includes social media information. And if it's a consumer reporting agency, it's going to be very restricted, if you will, and very calculated and carefully synchronized with what the Fair Credit Reporting Act would say with respect to reporting that information. But they're only going to be looking at certain things, it's a much smaller universe, whether its illegal activity or racist comments or explicit photos.

I mean, that's what a consumer reporting agency that would look at and provide a report that includes social media would look at, because what they want to factor out for their sake and for employers' sake are the discriminatory elements that one could see if, for instance whether it's religion or maybe -- certainly gender, that you would see if you were just an employer who is Googling it.

So, certainly, I think that's an area where FCRA provides a lot of protections for consumers, if an employer is, in fact, going to request a report that includes social media.

MR. SWIRE: Here's one distinction that hasn't been brought up in the panel. Under ECOA, you don't usually think of there being different loans to women or men. And, in fact, one big reason why the Equal Credit Opportunity Act exists was to correct for a history where married women didn't get their own credit history, it was just the husband's credit history. And divorced women turned out not to have a credit history and couldn't get a loan, once they were divorced.

So, in the credit area we don't expect there to be men's loans and women's loans, or black loans and white loans. That would be very -- we'd be extremely skeptical of that in a credit relationship. The shavers and cosmetics categories, although shavers, I believe, are used by both sexes, but --

MR. PEELER: Not the same ones.

MR. SWIRE: Well, I don't know the facts on that.

(Laughter.)

MR. SWIRE: Anyway, in cosmetics you can get into your own discussion. But I think for some universe where there does seem credit related and we have some -- there's some uncertainty about what sort of things are going to be credit related, it might turn out there's advertising that's directed more towards one sex or another, one national origin group or whatever.

And where one of these statutes applies, it doesn't mean that you can't, under the law, turn out to have a women-targeted ad or a men-targeted ad. If one of the discrimination statutes applies -- lending, housing, employment -- my understanding is then it's a business necessity defense. You get to do it because we have to do that in order to sell the cosmetics or whatever it is.

But there's a prior question of when are these statutes going to apply and once they do, you can have a defense of necessity, but then the company has to come forward and show the facts supporting that.

MS. WORTHMAN: Building a little bit more on the social media comment, what about employers who look at social media to determine hiring eligibilities? Or also in some lending context where people look at how many friends you have or who your friends are in determining whether or not you're eligible for credit.

MS. MIASKOFF: Right. Well, in just looking at employment, employers who look at social media as part of the screening of applicants, you know, frankly, it puts them, I would think, in a vulnerable position, vis-a-vis the EEO laws. Because, obviously, with many social media you take one glance at it and you learn, you know, a plethora of information about various protected statuses the person may have.

And once the employer has that information, if they deny the job to the individual or they deny the promotion or the training, and the person is trying to think, gee, why didn't I get this? And they happen to find out, perhaps, that social media was looked at. You know, it's -- they may well bring a charge to challenge it.

And so from an employer's perspective you really have to step back and think am I going to get something that's really, you know, related to job performance and worth my while here for taking that risk.

MS. MILLER: And I would also say with employment, in bringing it back to the FCRA, the biggest challenge with social media is just accuracy.

MS. MIASKOFF: Yes.

MS. MILLER: And -- so which is why consumer reporting agencies would just look at user generated content, as opposed to any content that's out there. And then the other question, of course, which is not so much FCRA, just as its terms of service or their privacy policy, is depending on how you capture that social media. A consumer reporting agency would need to look at just what's publicly available. You have to be careful not to go beyond the bounds of a company like a Google or a LinkedIn or Instagram's either privacy policy or terms of service and capture information that is in violation of either of those.

MR. SWIRE: A question on whether there could be another concern about social media being used in recruitment, for instance for employment, it may well be that people have a lot of friends who come from the same ethnic, racial, whatever background as themselves --

MS. MIASKOFF: Right, right.

MR. SWIRE: -- and so if you're trying to have diverse recruitment and it turns out you're sort of going down a path that's very dependent on one group, that could raise the EEO question as well.

MS. MIASKOFF: It does raise EEO questions. And the answer to it is that you have to recruit through many different sources and avenues and tools to sort of counter-balance that. I mean, there's also just an issue in terms of, you know, computer access, period. Smart phone access, which many more people have now, but still there are people who don't have it. And you certainly can't access as well some things on a smart phone as on a computer.

I'd also add with the social media, as you may be aware, there are many states now that have laws that prohibit employers from requiring people to give them their social media passwords to check it out. There was pending Federal legislation, but that has not gone anywhere. Though I've certainly heard stories that despite that legislation you have employers saying, now I'm going to turn my back so I don't get the password, but log in now and I want to see it.

MS. WORTHMAN: And in the credit context, Leonard, with the social media?

MR. CHANIN: So, I think you've got to divide between the marketing, based on that information, versus a customer. So there is nothing in Regulation B that prohibits use of the information, but I would be very careful because as was suggested before what's on that website, if you have gender, racial information, ethnicity, age and so forth. If you look at that and then there's going to be certainly an allegation or potential allegation that you've considered it, either -- certainly if you have an existing customer, in terms of that customer relationship, potentially with marketing if you have that, certainly there will be questions to follow.

So I’d be very careful about using it. Even though there’s nothing that directly prohibits use of social media, at least in the context of credit transactions.

MS. MIASKOFF: And I would just add, with social media and employment -- although I think probably the rule, rather than the exception, is people tend to have as their friends, they have people from similar backgrounds as themselves. I know, you know, sometimes I have a variety -- I’m just using myself as an example -- a variety of friends and as a result of that I get some very interesting suggestions from Facebook as to, you know, what group I might want to join or whatever, what publication I might want to follow. And were an employer to look at that, they could, you know, then draw conclusions about me.

So there’s really a lot of vulnerability for employers.

MR. EAGAN-VAN METER: Are current categories or protected groups under anti-discrimination and consumer protection laws sufficient? Panels 1 and 2 discussed victims of crime or domestic violence, as well as people with particular health statuses.

MS. MIASKOFF: I would jump in. I mean, I think basically, yes. In terms of health status, with the expanded definition of disability that came into effect in 2009, there are a lot of health statuses that are covered by the ADA now.

I think, you know, societally we may -- you know, the big categories that are covered are the ones that our society has had major, major problems with. And I think that's sort of an appropriate focus for these laws. In terms of abused women, I think possibly the gender -- gender could capture that, possibility disability in some ways.

MR. EAGAN-VAN METER: How effective are adverse action notices under ECOA at conveying an adverse credit decision, where that decision might be based on thousands of big data variables?

MR. CHANIN: I don't know who is taking that one.

MR. EAGAN-VAN METER: You have an audience member to thank for that.

MR. CHANIN: I'll stall and let them have time to think, but only because Katie has told me that we are going to discuss big data and NSA before we leave today.

MS. WORTHMAN: No comment.

MR. CHANIN: How can you have a discussion about big data without discussing the NSA?

I think the answer to that question is, at least to my knowledge, we don't know. Adverse action notices, you either have to give automatically the reason for the denial, they have to be specific, or the consumer has the right to get the specific reasons. They have to be pretty detailed, so if you use credit report information or any other information, you have to give enough information so that a typical consumer can understand exactly what it is.

So if the person has been late in making payments, if he or she has a charged off account, filed for bankruptcy, all of those sort of things have to be clearly communicated. What's not clear is -- or at least I'm not aware of any data that has studied, you know, what consumers do with that information. To the extent they can, are they able to correct the information moving forward, or how do they use that information.

It might be an interesting research topic, but I'm not aware of any data on that.

MS. MILLER: And I would just say under FCRA, I mean, certainly adverse action notices are built into the FCRA. It's an important part of it, whenever you're using a consumer report and it's for one of the purposes, whether it be employment or tenancy or credit, you have to provide the adverse action if any information from the report, including maybe if it’s credit information, is used in whole or in part to make an adverse decision.

And then take it one step further, you have employment and there is an additional pre-adverse action step that needs to be followed. If information in a report is going to be used adversely against an individual, they must be provided notice of that and a copy of the report and a summary of their rights. So certainly adverse action is built into FCRA.

MR. SWIRE: One other thing about adverse action notices is it’s not just whether that individual cures their problem. Another role of them is an enforcement regime overall, so if there’s an adverse action notice that might end up with an advocacy group or a plaintiff’s lawyer realizing there’s some practice that should be challenged and maybe a complaint to a regulator.

And if they’re not being issued the adverse action notices, that can get detected with the company and lead to enforcement.

So it’s part of an overall structure to detect things that might turn out to be troublesome and it’s not just the individual fixing their own credit.

MR. PEELER: And so, as historian of the panel, to put a little context on that discussion. You know, the adverse action, ECOA, Fair Credit Reporting Act structure, you know, created a dynamic where you have greatly expanded, where big data greatly expanded credit availability to consumers, made the decision making a whole lot more objective and built in some checks and balances.

So, you know, like one of the big challenges that was alluded to this morning for big data in the credit area is to expand that model to, you know, consumers that currently can't be credit-scored.

MS. WORTHMAN: Lee, you have the last word on that.

I'd like to thank all of our panelists for the discussion, it's been very informative.

(Applause.)

MS. WORTHMAN: And we are now going to take a break and return at 3:15. Thank you.

(Whereupon, there was a brief recess.)

PANEL 4: CONSIDERATIONS ON THE PATH FORWARD

MR. OLSEN: Thanks, everyone, for joining us for the final panel. We're here talking about big data. A lot of people talk about leaving digital footprints. Somebody left physical evidence of their person in the lady's room, some reading glasses. They're up here to be claimed.

I know there's no coffee allowed in here, which I think is sort of a disaster for the last panel of the day. So, we screwed everything up, didn't we? We have screwed everything up, and we haven't even started.

MR. CALABRESE: I blame the FTC.

MR. OLSEN: All right, this panel is on paths forward. I have a very distinguished group of panelists here with me. It's going to be a challenge for all of us because a number of panelists earlier in the day discussed steps forward. So, this panel is challenged to come up with something new and different for the last panel, but I'm sure they're up to the task.

Just quick introductions. I should have borrowed the reading glasses that I just had. To my left is Chris Calabrese who is the legislative counsel for privacy related issues in the ACLU's Washington office, where his portfolio includes internet privacy and new surveillance technologies.

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

Jeanette Fitzgerald, next to Dan, is general counsel and chief privacy officer for Epsilon, where she leads the government affairs legislative and regulatory initiatives related to data protection and privacy.

Jeremy Gillula, did I pronounce that right?

MR. GILLULA: Yes, you got it.

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

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

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

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

MR. GILLULA: We were told there would be no quizzes.

MR. OLSEN: So, I’m going to start with sort of a McLaughlin Group style question. There’s been a lot of discussion today about practices that are occurring and could occur. And there's been discussion about the legal landscape and the regulatory landscape.

I'd like to ask each panelist for a yes or no answer to the following question. You can say it depends, but that's really cheating. So, I wouldn't go with that. Do you agree that there are currently uses of data, or potential uses, that are harmful that are not addressed by the current legal or regulatory landscape?

Chris?

MR. CALABRESE: Yes.

MR. OLSEN: Dan?

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

MS. FITZGERALD: No.

MR. GILLULA: Definitely.

MR. SPADEA: Gun to my head, no.

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

MR. OLSEN: Okay, it sounds like we've got a mix on the panel. I think before we get too much into specifics about how we might move forward, it might behoove us to flesh out a little bit of the answers that have been given to the simple question there.

I would ask each panelist to talk about whether there are legal gaps or market failures that are not being addressed in the first instance. I'll just start with Chris.

MR. CALABRESE: Sure. So, just to give a frame for this, data is not bad. It's not good either. It just is. It's a fact of the environment, so it reflects existing disparities in our society. You know, we see a lot of money in this country that is distributed along racial lines. So, we are going to see those distinctions.

I believe the wealth gap in this country -- white households now have approximately 20 times the average household wealth of black households. So, the data is going to reflect that. So, our job here is to make sure that big data does not exacerbate it and then, ideally, hopefully down the road, can help to close it. But let's start by not exacerbating it.

So, potential regulatory gaps, I am very comfortable saying that there are regulatory gaps, and I'll give you a couple. One of the major ways that big data and data is combined today is in background checks. So, if you want to see whether somebody's got a criminal background or not, and I know it's covered, is you do this background check. Lots of public data sources are checked.

We see Chris Calabrese's criminal record in Texas. That is not true, by the way. So, Chris Calabrese probably doesn't get a job. Well, we've seen lots and lots of examples where there are multiple Chris Calabreses, and there are, and those are mixed up.

Well, I see a great and classic example for a market failure here because the customer is not Chris Calabrese. The customer is the company, and he or she, it may be willing to deal with a certain level of error if it improves their bottom line which is not to hire somebody with a criminal record. They may be willing to accept a certain amount of data problem in order to deal with that larger problem.

Similarly, I have a product that detects fraud, right. If I'm a big bank, I'm really excited if I cut my fraud down by 40 percent. If I have two or three percent of people who aren't able to get a product or have to jump through more hoops to get a product, that's fine, because that's not really what I'm worried about, right. My desire is to reduce fraud. I'm willing to accept a certain amount of error to do that. If some people don't get products, you know, that's too bad, but again, the market isn't going to fix that.

So, I'll just leave it at those two.

MR. OLSEN: How about you, Dan?

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

I think that's what we have to talk about when we talk about regulatory gaps. It's not enough to say there might be a problem. The reason this matters is because there is so many opportunities to use big data. That was, you know, part of what the first panel talked about.

So, when we're talking about regulatory actions, we know there can be unintended consequences. There's always unintended consequences with any action. So, we have to be asking, you know, can in this case the FTC make a good cost benefit analysis of any type of proposed action, any type of proposed intervention. You have to know what the costs are. You have to know what the savings will be.

But just to also pick up on something that Chris had said. You had said, you know, data just is. I guess it depends on what your definition of “is” is here. But I think that’s actually the wrong approach. We just came off the legal panel, so, you know, I’ll throw it back to Lawrence Lessig and his famous line about code is law. In this case, data is law.

Data isn’t natural. It’s something that’s created. We have to think about how it’s created and the implications of this creation. Part of us doing that helps solve some of these types of problems that we confront. That’s not a regulatory solution; that’s a technology solution.

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

Jeanette, how about you?

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

So, I said no because I think there's a lot of self-regulation that I think already exists. The DMAs out there, the IAB, and there are several other of those As and Bs and all those other groups that have all self-regulatory guidelines.

I know for the DMA they will enforce those guidelines among their members. If they hear about somebody who is not a member, they will go talk to them and try and get them to act in what is considered an ethical manner among that group. If they then still find that there's a problem, they've been known to turn those companies over to the FTC so they can look at them further.

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

MR. CALABRESE: I have more.

MS. FITZGERALD: Okay, good.

MR. OLSEN: How about you, Jeremy.

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

If you're talking about classifying every person in this country, that means you're wrong six million times, you know, or more, if you've got a two or three percent error rate. That's a lot of people that your automated decision making based on big data could be harming.

I think it's a different thing when you're doing a scientific study using big data. You're looking at a lot of data about health and trying to make a determination about, you know, what causes this disease. It's a different thing when you're testing it on people. It's tough to tell when you actually have a false positive or a false negative.

So, I think from a technology perspective, we need to make sure that the underlying technology is really working as we think it should.

MR. OLSEN: Okay, thank you.

Michael.

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

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

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

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

MR. WOLF: So, I think the other Chris really hit it on the head when he says that data is neither, per se, good nor bad. I think that really ought to be the guiding light here, because we have seen that there's enormous potential for good with the use of big data.

I think we're going to get into this a little bit later, but thanks to Mark MacCarthy for previewing the study that the Anti-Defamation League and the Future Privacy Forum did on the beneficial uses of big data to identify discrimination and therefore to come up with remedies for it and also big data as a tool to fight discrimination.

So, this is the baby and the bath water theory of regulation that I typically espouse. We need to be careful when we're identifying potential problems or even real problems in regulating in a way that throws the baby out with the bath water, and that might have the unintended adverse consequence of inhibiting the positive uses of big data. I know the FTC has that in mind.

It's been, I think, a balanced day, and I'm hoping this panel will continue to be a balanced discussion of that issue.

MR. OLSEN: Thanks, Chris. I wanted to follow up on two different comments, one that Jeanette made about self reg. I just wanted to pose a question following on Latanya's presentation and ask about the Omega Psi Phi example, the ads being shown related to getting your arrest record, hiring a criminal lawyer, perhaps getting less advantageous credit card offers.

Where does the self-reg fit in that scheme? What are you -- given your position in the industry -- what's your explanation for that particular scenario, understanding you don't have any of the facts other than what was observed?

MS. FITZGERALD: Exactly. I have no facts, but I --

MR. OLSEN: But I think you can see the web site has a particular demographic, and there are particular ads being delivered.

MS. FITZGERALD: Sure.

MR. OLSEN: Something is going on in the machine somewhere. Where does the self reg kick in there? Is that anecdote potentially harmful, troubling, concerning to you? Is there a role for self reg there that would address that scenario?

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

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

But, in my mind, self-regulation says if you're part of this industry, if you want to be part of these groups, we are going to use the data in a responsible way. We are not going to try and violate anybody's rights. But we're only going to use it for marketing purposes, because in the end, it’s an advertisement.

It’s the same thing you get on TV. It’s an advertisement. You can either take it or leave it. If you want a different offer, go to a different bank. You don’t like that bank or you want to see if the bank has something else to offer, go talk to the bank.

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

MR. OLSEN: Okay. So, given that, it was simply the delivery of ads? It doesn’t present an issue that the self-reg guidelines or --

MS. FITZGERALD: Yes.

MR. OLSEN: Does anyone want to address that or comment on it before we move on?

MR. CALABRESE: I guess I’m a little skeptical. I’m not sure this is directed at self reg, but I’m a little skeptical of the idea of it’s just marketing, actually gets you all the way to where you want to go. One of the things, Commissioner Brill’s concurrence to the recent data broker report, she talks about the use of aggregated credit scores.

I’m not entirely sure I understood what that actually means, given what I understand how a credit score works. But the idea is that you’re averaging credit reports in 5 to 10 households in a specific geographic area. I presume that you are using those for things like marketing and determining what kind of ads you are going to share with people.

To my mind, if, given the segmented and personalized nature of today's internet, if we are replicating the geographic segregation in our society and people are seeing, based on what neighborhoods they are in, different types of ads and offers, that is problematic, full stop.

Even if they can go to another bank, if all they are seeing are the crappy credit card offers again and again -- and maybe because they, you know, are the kind of people who go and get the only advertisement they see, they don't know to go to another bank. They think that's the bank, those are the offers they get.

So, to my mind, that kind of stuff is where a market failure exists, where the CFPB should push harder to see if those offers are actually dissuading people from getting credit or if they are ending up with worse credit offers because of them. So, that's, I think, an area to push.

I just don't think the industry self-regulatory model can fix that. Now, aggressive regulation may be able to, but I just don't think that saying it's just marketing is sort of enough to answer those kinds of criticisms.

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

MS. FITZGERALD: Am I allowed to?

MR. OLSEN: It's totally up to you. Yes, I invite you to.

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

MR. CALABRESE: True.

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

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

MR. CALABRESE: I'm sorry, it's just not. If you were looking at the credit scores of different people in the apartment and aggregating them, which means if Dan has got much better credit, and Chris has got much better credit, and I've got worse credit, and I am bringing them down, and they are getting worse offers, that is not the same thing. It's not the same thing as where they are or whether they can mow a lawn. It's different.

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

Sorry I interrupted you. I apologize.

MS. FITZGERALD: It's okay.

MR. OLSEN: Dan, did you want to say something?

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

MR. CALABRESE: Separate apartments.

MR. CASTRO: Separate apartments. Just to be clear.

MR. CALABRESE: I'm just too old to share.

MR. CASTRO: I was thinking three bedrooms.

MR. CALABRESE: Right.

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

MR. CASTRO: So, if we're sharing an apartment building --

MR. CALABRESE: Right. That was actually my point.

MR. CASTRO: So, but the point here is, though, what is going to happen over time, right. Because the question is, you know, if I'm getting worse offers or you're getting worse offers, then there's a market opportunity there, right. So, there's an opportunity for another company to come in and steal this business. That's something good. I mean, that's the kind of innovation we want to see. So, that's not a market failure; that's a market opportunity.

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

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

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

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

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

MR. SPADEA: I would change the question slightly about the uncertainty part to call it, and I think it was pointed out earlier, a very, you know, nascent industry. It's brand spanking new, really. I think everybody is trying to feel their way along about the risk benefit, what's ethically appropriate.

I think we need to hear more from economists as to, you know, the risk benefit analysis. What is the economic impact that new regulation may have? Will it promote trust unless there's a benefit there? What is the economic benefit or loss to consumers who have to spend time trying to remedy inaccurate information?

What's the drain there from time, money? For middle class families, it's not as much, but when you think of low-income families that are wage earners taking a day or two off to deal with something like this, that could have a very dramatic impact on them.

I think we need to hear more from ethicists and try to look at, you know, what can be taken in from -- institutional review boards were mentioned earlier. Is there some good practice there that can be pulled in? How do we look at harm in those situations? More from ethicists in general to help us figure out, you know, what’s right and wrong. Should there be -- harm include things other than just economic harm as well?

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

MR. OLSEN: I think you teed up Chris’s --

MR. WOLF: Well, actually, before I get to the FPF ADL report --

MR. SPADEA: I did that on purpose.

MR. WOLF: -- I just wanted to add to what Michael said, because recently the Berkeley Information School folks asked 40 thought leaders what big data was, and there was 40 different answers. I think the one slightly negative comment I will make about some of the discussion today is we’re painting with an awfully broad brush in talking about big data and talking about harm as the same thing in all contexts.

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

So, to help do that, the Future of Privacy Forum just published something called Benefit Risk Analysis for Big Data Projects, which tries to provide a framework that can be used across the 40 or more instances of big data and the many potential uses and harms and really moves privacy impact assessments forward to talk about data benefit analysis.

So, I commend folks here and those watching to take a look at some of the work that my colleagues, Jules Polonetsky, Omer Tene, and Joe Jerome have done.

MR. CALABRESE: Can I offer a countervailing report? I've read Chris's report. I think it's very good. Everyone should also then read David Robinson's report, which I think also tackles very specific and concrete examples and I believe takes a little bit of a more critical view of some of the areas.

Everybody in the civil rights community agrees that data is a good thing and can help things. But David talks a little bit about some of the complexity of algorithms. So, it's David Robinson's report. Sorry, I just thought I'd balance it.

MR. OLSEN: That's fine.

MR. CALABRESE: They're available as a box set.

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

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

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

MR. OLSEN: Okay. How public are those analyses?

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

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

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

MR. OLSEN: Jeanette, what are you seeing? Is that something that Epsilon does, this sort of cost benefit analysis?

MS. FITZGERALD: Absolutely.

MR. OLSEN: And just to key off my earlier question, I think there was discussion at one of the earlier panels, I can't remember which one, but I think it was danah boyd who said, you know, there's a lot of public uses of data and data sets that are very transparent, how the data is crunched, how the data is analyzed, what the results are. All of that is made public. There is no similar transparency on the commercial side.

I think the cost benefit analysis, the benefit risk analysis, sounds like something responsible companies should be doing today. The question is, how does anyone get any sort of comfort that the analysis is either not affected by concerns we would care about or the results aren't unfairly impacting someone? How do we get over that transparency hurdle?

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

So, it’s evolving now. As Chris Wolf said, the IAPP, it sells out all the time. There’s always people that are -- I see the same group all the time, but we’re all talking about new issues as it evolves. If we, as a company, my company, for example, Epsilon, decided to tell everybody exactly how we did a risk benefit analysis, that would be giving up trade secrets. We’re not going to do that. Other companies are going to feel the same way. That’s part of our “special sauce” to make it.

Now, that doesn’t mean that if there was an impact that somebody felt was discriminatory, that somebody is not going to come back to us and say, you know, there’s a problem here. Then, what happens? Our name shows up somewhere. We don’t want our name to show up. So, there’s a lot of good reasons why we’re very careful about those things.

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

MR. OLSEN: Okay.

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

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

MR. CALABRESE: I would like to. I think we’re sort of woefully inadequate when it comes to transparency right now, so I’ll pick on the data brokers just because there was a recent report. There’s a fair amount we know about exactly what their practices are.

In the recent FTC report, I believe it was Acxiom, they said Acxiom had something like 1,500 or more than that data points on every consumer. I went and looked at my Acxiom profile. There’s nowhere near 1,500 data points about me, nor, and I think more importantly, is there anything about how they’re being used, like sort of what score or assessment am I being offered.

Am I being grouped as an urban scrambler? Am I, like, a vulnerable consumer? In my personal life, I'm very vulnerable. But until I know those assessments, and I get that it's special sauce, but the individual consumer should be able to know if they're being targeted or there's an assumption being made about their health, their finances. So, that's the kind of transparency I'd like to see more of.

MS. FITZGERALD: So, let me throw one thing out in response to what you're saying that Acxiom does. Epsilon, too, has a site, a section within our site, I don't know whether it's called a microsite or not, because I'm never even sure what those are, where a consumer can go and we give them education about how their data is used for marketing purposes, places they can go with the DMA, places they can go to the FTC to get further information. They can go to a couple different other sites that will give them lots of information about how it's used. I don't know how many people are actually going to go read all that stuff. It takes a long time to slog through it because we all did before we put it up there.

But the other thing it does is it says, look, if you want some information about you, if you want to know what kind of information we have about you and what group we put you in, you have to send us something because at this moment we haven't figured out the right way to do it online. I've got a couple ideas, but I don't like what I've seen so far, because we don't have credit cards, we don't have driver's licenses, we don't have social security, and I'm not going to ask for those to then give you a report. That seems like counterproductive.

But we can give you a report, and I should have brought it. I'm sorry I didn't. But basically, it says your name, your address, whether you have kids in your house, whether you own your home, and then some of your interests. So, we like to be outdoors. Yes, we do hunt because we're in Texas. Fish.

And then, it says at the bottom you're in this group, and it says these are the basic characteristics of this group. The salary is about this, you buy books and magazines, you shop online. I can't remember what the other ones are. But, I mean, that's available. You can go get that and we'll show you. Frankly, after you read that, if you're really still worried, we'll opt you out.

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

MR. SPADEA: If I could, just briefly, I don't know if the answer is more transparency but perhaps better transparency. We've all heard about the studies where, you know, it will take you 29 years or something to read all the privacy policies. It's not because consumers can't do it if they want to; it's just, who has the time to do that.

If we inundate consumers with descriptions of the technologies and the business processes and all the data flows, they're not going to read past the second privacy policy. Think about when you buy a new computer and reinstall your software and reupdate the stuff from the cloud. You're not reading all that. I don't read all that.

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

You want to deliver the critical information at the moment in time and that's that. Perhaps a better approach may be, which we all do in some ways, you know, risk rate the data for that data that's most sensitive, that might have the most potential impact. There's a higher, you know, notice requirement there. But just a blanket across the board, dump tons of more information on consumers, I can't see that as protecting consumers. In fact, it may put them at more risk.

MR. OLSEN: Michael, I don't disagree with that. I think in terms of transparency, there are a variety of ways to deliver that. I think what I was contemplating is some mechanism for companies -- I think Peter Swire alluded to this in the last panel -- if there is an unfair practice, if there's unfair marketing going on, you could foresee a scenario where the business has a justification for why it engaged in the particular marketing campaign.

It's not necessarily a justification for consumers; it could be a justification for a self regulatory governing body like DMA. It could be a justification for regulators. It's not necessarily giving a notice to consumers every time they receive an ad that says this ad was delivered based on the following 15,000 analytical data points. I don’t see that as being particularly helpful.

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

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

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

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

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

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

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

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

This gets to a really important point about whether, you know, the discrimination or harm that you're positing here is something that's intentional or unintentional, which is something that the first panel talked a lot about. How you address those might be different, so you need to think about which type of problem you're trying to address.

If it's, you know, unintentional harm, you really do have to address a lot of that through data analysis. Nobody is intending to do it, so you have to make sure you have smart data scientists doing things consciously, but also that you're able to evaluate outcomes. If it's intentional, then you have a human problem, and maybe you address that differently.

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

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

What you really want to get to, I think, is where getting access to the data is really easy and it's all the intelligence and innovation that you build on top of that that's hard. So, you want to promote that. So, I think part of that is by allowing trade secrets to exist, is by allowing, you know, intellectual property to be protected here.

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

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

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

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

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

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

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

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

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

MR. OLSEN: Right, and those are great examples. I think they all serve to really demonstrate that big data has tremendous societal value.

Let me first jump in here for a second. I think what we're talking about, I think, Chris, maybe you and others on the panel, are teeing up the scenario of regulation versus no regulation, law versus no law. That's certainly an option. Should we have another law?

Should we recommend legislation? Should congress pass legislation? That's an option.

We all know how challenging it is for something to come out of congress. So, let's talk about best practices. Are there practices that companies can engage in that would measure, or cabin, or restrict, or evaluate potentially harmfully uses that they're not going to impact potentially beneficial big data uses?

I mean, if you have a data ethicist or a chief fairness officer, you know, that person and the evaluation that that person undertakes before a new program is rolled out, it's not going to curtail the benefits of big data.

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

MR. OLSEN: Okay. How do we come up with a strong ethical code?

MR. WOLF: So, I have lots of booklets to waive around, but I come back to the original one, which is the benefit risk analysis for data projects. Having a framework and a methodology and a discipline within an organization is absolutely essential.

Now, is it like a traditional IRB? Maybe not, but Professor Ryan Calo has said that it's certainly something to think about as a way to if there is a gap, fill the gap. But even if there's not a gap, to avoid adverse consequences.

MR. OLSEN: Jeremy, you wanted to jump in?

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

MR. WOLF: That then benefitted individuals to allow them to have an education free of discrimination or healthcare free of discrimination.

Mr. GILLULA: Sure, absolutely. But, you know, that's called science. Just that we have more data or we collect more data, we can find more things. But what we're talking about, a lot of the harms, are harms where data about a person is then used to make a decision about that person, not we found some trend and then we adjusted our methods.

We found some trend and then Google decided, you know, yes, we need to change our hiring practices. Google looked at your data when, you know, they were deciding whether or not to bring you into an interview and based on the data decided not to bring you in.

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

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

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

MR. CALABRESE: I'm not assuming.

MR. WOLF: No one has said that today.

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

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

MR. CALABRESE: What I'm saying --

MR. SPADEA: Let me just finish one thing. I was going to add, though, I do agree to your point where there is harm, you know, there should be some type of remedy. We shouldn't just leave consumers floundering. The question was, do we need a law or not. I think what I'm trying to say is that the evidence to say that we need legislation now is not there.

As this industry continues to develop and we have more information about harms, about benefits, we need to continue having this discussion, and there may come a point where we do need further regulation or legislation. But we need more information. We should start with the least interventionist approach. If that doesn't work, we rachet up the intervention.

MR. OLSEN: I guess the question I would ask before I turn it over to Dan goes back to the transparency question. If there is no transparency about how the data is being used, then how do we get to the we have more information point in order to make a decision? It may be that companies internally know how they're using the data, but they're the only ones who know that.

So, Dan, you wanted to --

MR. CASTRO: I think this would be a good point to talk about this paper that we released last week.

MR. CALABRESE: Everyone else did.

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

MR. CASTRO: Since you brought it up --

MR. SPADEA: It's the afternoon panel.

MR. CASTRO: We released a paper called the Rise of Data Poverty in America. This gets to what Christopher was talking about, which is, you know, what are the individual benefits and risks. So, the point of this paper, we went through and we talked about specific benefits that individuals are seeing in areas like healthcare, education, and financial services.

We also talked about the challenges that we've had in the past, both in terms of the digital divide and how that might translate into a data divide, as well as challenges that we've had in small data sets. So, you know, the best example of this is in healthcare where historically, if you look at, for example, clinical trials, minorities and women have been underrepresented in this data. Just as when we're talking about big data, decisions are made from big data, decisions are made from small data as well.

So, decisions have been made in the past, for example, by the FDA about what drugs and treatments were safe and effective. It turned out that, of course, once they release it to the full population, that population didn't match up with the clinical trial population. Some things were unsafe.

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

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

The policy answer to that type of problem of a data divide is to say how do we make sure that disadvantaged populations also have data available about them, that they can share in these benefits. When you look at it, it's very clear that there are huge, you know, economic, educational, health benefits. We want to make sure all of these groups can share in that.

MR. OLSEN: Yeah, that brings up a good point. It reminds me of the street bump example where data was being collected about road conditions from smart phones. There was a question about how broadly representative different communities were in that data collection.

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

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

MR. OLSEN: Correct.

MR. CASTRO: Hopefully, you know, we're doing some of that today.

MR. OLSEN: Michael.

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

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

Well, you know you need to have something else in place for those neighborhoods. So, you take the money that you save from pulling DPW people on pothole patrol, or whatever they do, you take some of them, you take half of them and you just take that money savings and you put it someplace else. You take the other half and just throw them right into the neighborhood that doesn't have that saturation. At the end of the day, you get to the same place. You get there more cheaply. Everybody is happy.

So, you can, where you know where a problem like that exists -- and the key thing is the governance that Chris talked about earlier. There should be a process to spot those risks. The ethics piece comes in where, okay, we now need to fix it. We can't just let that harm sit out there. But we can still roll forward with the application and, you know, with a private company, generate revenue and service the consumer.

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

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

MR. WOLF: So, maybe I can start. I referenced the fact there are at least 40 different definitions of big data, but there’s one kind of fundamental understanding. It relies on volume, variety, and velocity of data that leads to unexpected discoveries.

So, how do you provide notice at the time of collection to allow consumers to make choices about discoveries that you don’t know will happen? That’s sort of conceptually one of the problems I have with this idea of a collection limitation.

But I think a more practical issue is one that I think Pam Dixon very candidly acknowledged, is that there are huge data sets already out there, structured and unstructured, data exhaust, as she referred to it. Even if we are able to minimize data collection or to provide options that put limits on the collection, we're still dealing with huge issues of use.

As we discussed here today on this panel and others, consumers aren't simply going to take advantage of the transparency options and make the choices that perhaps we think they ought to. There has to be someone responsible in the ecosystem. That's why, you know, we urge the governance model and the focus on use without rejecting the FIPP of collection, but without unduly placing emphasis on it.

MR. OLSEN: Anyone else want to comment on this?

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

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

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

MR. OLSEN: Anyone else on this particular point?

MR. CASTRO: Well, just to a couple things that have been mentioned here. I think we have to separate between harms to an individual as in, you know, I’m paying more than I’m paying today versus, you know, I’m paying more than I’m paying today because someone has something wrong about me, right.

Like the e-Verify example, if I can’t work because the government, you know, fundamentally has something wrong about me, regardless of the law itself, you know, that’s a different problem than if my insurance company charges me more because I speed a lot and now they know about it.

We need to separate out those types of –

MR. WOLF: One is a harm and one is actually an improvement because there's actually a benefit to somebody else.

MR. CALABRESE: There's an information asymmetry here, right. If I know that you are wealthy and you are more likely to come into my store if I give you a really robust coupon, say a $15 coupon, but if I know you're income, I don't have to give you a $15 coupon. I can get you in the store with a $3 coupon.

Now, we can argue about whether that's a genuine harm or not and you can shop somewhere else, but the fact is that you know something about me and you're using that to provide a differential in something that I would value.

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

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

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

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

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

So, should consumers consider other options that exist or should we consider other options from a technological standpoint? Should we push for data tagging, for example, that would identify the provenance of data elements, or are we beyond that? Or, should we consider, you know, random identifiers that would mask your identity as you navigate the web so that you appear to be a new person every time you visit a particular web site? Or should we entrust our data to a third party with a permissions scheme? Are there measures consumers can take or companies can deliver that would mitigate the risk that the data would be used in harmful ways?

MR. WOLF: Unlike your first question, which I refuse to answer, which is a yes or no question, the answer to all of your questions is yes. I think technology does have potentially a very significant role here to play to provide exactly those kinds of protections, exactly those kinds of options. You didn't say de-identification specifically, but I think that was implicit in your question about random identifiers. So, you know, I think there's great hope in technologists. They've certainly gotten a lot richer than lawyers.

MR. OLSEN: Even you, Chris?

MR. WOLF: Hey, that's private information.

MR. CALABRESE: I've seen those.

MR. OLSEN: Anyone else want to address it?

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

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

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

MR. OLSEN: Anyone hear that who has the purse strings?

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

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

I'll start here, and we'll move down.

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

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

MR. OLSEN: Thank you.

Dan.

MR. CASTRO: So, I'd say, you know, I think this is definitely the start of the conversation. We need many more voices here. I think it's interesting. Today's workshop has been fantastic, but, you know, across town, there's a predictive analytics government conference going on with some of the best predictive analytics data scientists in the country, and they're not in the room. So, you know, we need them here. They certainly should be part of the conversation.

I guess my recommendation here is that, you know, we really need to be thinking about the benefits here. To me, if you care about discrimination, if you're worried about healthcare or improving education for our kids, the biggest risk is not how data is being used; it's that we won't use it enough. We need to figure out a regulatory environment and policy recommendations to help encourage more use of data.

MR. OLSEN: Jeanette.

MS. FITZGERALD: So, I would say that we need to spend time figuring out the best way to educate consumers about the data that's being used on them, about them. It's not just how it's being used, but we also need to teach them that they can talk to the companies that have their data. They can ask questions. Those people will help them understand what information they have and how it's being used.

I would encourage any other company that's been thrown in the data broker realm that they, too, think about ways that they can show consumers the information they hold on them and how it's being used, what category people fall in.

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

Right now, if you try to go and find this stuff, it feels like diving into a deep and shadowy world. That may not be what they mean it to be, but that's what it feels like. I realize a lot of this is trade secrets, secret sauce, but even just sort of giving general ideas to consumers I think would be a huge benefit.

The other recommendation that I would make is actually towards the FTC and really anyone sort of observing this space. Look closer, look past the height. I'm going to reiterate this point that I said earlier because I don't think it was adequately addressed, that a lot of the benefits that people tout about big data are benefits that come from analyzing and learning things about a population.

For every, you know, 10 benefits of big data I hear about that, I hear maybe 1 about how individualized targeting did big data help people. It's that individualized targeting where I think a lot of the harm is. I don't think there's a lot of harm in learning about, you know, hey, look, these types of students need help or these interviews are harming people. It's when decisions affect individual people's lives that I think we need to start thinking about.

MR. OLSEN: Michael.

MR. SPADEA: I would urge companies to develop enterprise-wide risk programs. As part of that, have a data risk framework. I think you could just simply read all the papers that have been discussed or otherwise provided today to come up with a list of the potential risks.

You make determinations about what apply to your organization, make determinations about where your risk appetite is, and then put controls in place. While I guess some of the questions are difficult, a lot of it is not rocket science. Come up with the risks. Put the controls in place. Test against them. Have good governance in place.

I would say to everybody, we need to have a discussion about harm. I think that's central to how we move on from here. To the FTC and perhaps all regulators that play in this space, these workshops are great. It's been mentioned that we need to bring in some more economists and data ethicists and scientists and so on. So, everything just moved so quickly. It's like we should schedule the next big data workshop a year from now, schedule it now and get it done.

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

MR. WOLF: Is that another dig at lawyers?

MR. SPADEA: No. My lawyers are my best friends.

MR. WOLF: So, I’m hoping that just as privacy by design is entered into the lexicon of all privacy professionals, that data benefit analysis or benefit risk analysis with respect to the use of big data will also become something that’s reflexive and something that every privacy professional talks about.

I think that will avoid a problem I see with Jeremy’s analysis of focusing on who benefits. If you put rigid one size fits all restrictions on the collection and use of data, you’re not going to have benefits for anybody.

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

(Applause.)

MR. OLSEN: Jessica Rich, Director of the Bureau of Consumer Protection, is going to give closing remarks.

MS. RICH: So, good afternoon. Many of you are still here, I see. It’s great. We’ve had a really great day of discussion and debate regarding consumer protection issues surrounding big data and, in particular, its potential impact on certain consumer groups.

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

First, I want to thank the team, many of whom are sitting over there, that put together this terrific event: Tiffany George, Katherine Armstrong, and Chris Olsen here from the Division of Privacy and Identity Protection; Katie Worthman, Patrick Eagan-Van Meter, and Malini Mithal from our Division of Financial Practices; and Jessica Skretch and Lesley Fair from our Division of Consumer and Business Education.

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

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

Today we learned about many beneficial uses of big data. For example, we heard case studies about how big data can help fight discrimination, predict the risk of homelessness, increase diversity in the workplace, help ensure certain populations are getting the healthcare they need, and actually empower traditionally vulnerable populations.

But we also discussed the risk that big data can lead to selective opportunities, stigmatization, and discrimination. For example, Latanya Sweeney presented some interesting preliminary questions about how big data may impact the ads that visitors to certain web sites see based on the presumed race of the visitor.

Solon Barocas discussed the ways in which existing patterns of discrimination inherent biases present in the use of little data, such as the categorization of consumers based on their likelihood to buy can be replicated with potentially greater scope or scale in the use of bigger data.

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

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

We also discussed what happens when certain populations don't have the same sort of access to technology as other consumers. Will inequalities result from this lack of collection and use of data that could otherwise provide benefits to these populations?

We began, but hardly finished, discussing the overarching question that was the basis for this workshop: how will all of these new and evolving practices impact certain populations, and what steps can and should businesses take to make sure particular groups are not disproportionately or negatively affected?

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

To that end, our chairwoman this morning outlined three steps for moving forward, which I'll emphasize as my parting message. Actually, these three steps, or themes, or whatever you want to call them, were echoed down the line in this last panel.

First, as a law enforcement agency, the FTC will work to identify areas where big data practices violate the laws currently on the books that we enforce, including the FTC Act, the Fair Credit Reporting Act, and the Equal Credit Opportunity Act, and will bring enforcement actions where appropriate.

Second, we will continue our efforts to examine and raise awareness about the consumer protection and concerns surrounding big data through speeches, consumer and business education, which we certainly need to do more of, and potentially follow-up events or a report on this workshop.

And third, we will encourage businesses to design their analytical systems with an eye to the concerns that we've discussed here, avoiding bias or disparate adverse impact on particular populations of consumers.

Finally, I do want to mention that our comment period will be open until October 15th. Please don't be shy. Please comment if you have something to say. You can file comments electronically or by paper. The details are on our web site.

With that, let me just thank everyone for coming. Have a great evening.

(Whereupon, the proceeding was concluded.)

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: P145406
CASE TITLE: BIG DATA: A TOOL FOR INCLUSION OR EXCLUSION
DATE: SEPTEMBER 15, 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: SEPTEMBER 22, 2014

JENNIFER METCALF

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.

ELIZABETH M. FARRELL