FTC Big Data: A Tool for Inclusion or Exclusion? Workshop
September 15, 2014
Segment 3
Transcript

TIFFANY GEORGE: Good morning again. For those of you may missed the beginning, my name is Tiffany George, and I'm 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'll 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 the 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?

MARK MACCARTHY: Let me jump in. Is the mic on? Can you all hear me? 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. A 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% of Hispanics and African Americans could not be scored by traditional data systems while only 24% 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% 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 it before. But I wanted to put that one on the table as well. Cognitive computing in health care-- IBM has a version of its Watson computer that functions as an oncology diagnosis and treatment adviser. It's in use today at Memorial Sloan Kettering and M.D. Anderson. Mayo Clinic is using it to select subjects for clinical trials. So how does this help the underserved?

Well, there are shortages of specialty providers in hospitals all over the country. Some 50% to 60% 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 and 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 in the last panel, the use of 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 these, I'm sure, as we go on.

ALESSANDRO ACQUISTI: I'l do some marketing as well like Mark just did. [INAUDIBLE] is an economist at Duke, and [? Leo ?] [? 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 with them, 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 big data will increase the economic pie, will lead to more economic growth benefiting everyone, a win-win. And to what extent, instead, we'll 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 literature 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 HIE because it decreases privacy concerns and certainty that firms, health organizations, are [INAUDIBLE] in terms of how to use the data. In terms of the allocative effect, we find evidence of course that both privacy and lack of privacy affect the winners and losers. Sometimes, there's a 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 found is that even when candidates are identical in educational and professional background, there is an impact on personal information, protected [INAUDIBLE] such as religious affiliation or sexual orientation in how employers made 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 that, as she pointed out, big data will probably have both positive and negative economic 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.

CYNTHIA DWORK: Can I jump in here? So 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 Dana's 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?

PAMELA 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 really thought about this issue an awful lot for a lot of years now. And earlier this year, Bob Gelman 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 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, quote unquote, 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 questions 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 with 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.

NICOL TURNER-LEE: Thank you to the FTC for having me here in 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 things that I wanted 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 it's going to get better?

And I want to say yes, yes, and yes. Every day, we get-- I'm sure we 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 portfolio that, in some way, has an impact. And for social scientists like myself, who my own plug is I'm working on 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 US 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 analytics and data science?

Clearly, at MMTC, we represent underserved communities, particularly minorities and other vulnerable populations. Data analytics can certainly generate a social and community benefit. When I think about health care 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 that or educational outputs of value. Big data can in some way actually help us solve social problems related to health disparities, educational disparities, 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 economically depressed but at the same time create healthier communities and more sustainable communities. All that is great.

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 and minority kids. Again, that's all great. But as I said on our panel earlier or last week, it was Mark who 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 them 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 the social benefit without having a subsequent harm on those communities that are contributing to this?

And we've seen particularly with the FTC examples where some of those-- and I'm sure we're going to talk about it more in 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.

TIFFANY GEORGE: Stuart, I'm sure you have something you want to say?

STUART PRATT: Yeah. So I was invited late to this panel. So I missed the conference call. And Manisha called me and said, sure, 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 notes. 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. You were on the first one, and taking good notes.

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 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 someday, we're 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 are, as the Consumer Data Industry Association, we're much more focused on risk management. So 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 do 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 the CoA 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 first panel. But those laws are out there today. And I do think that that's part of the analysis going forward. 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 decision, a lender's decision to make what we'll call a risk based offer of credit. And of course, we 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.

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 that safety and soundness is a whole lot more important than 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 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%-- I see Nicol leaning in towards me like this, but I don't know we're 100% 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 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 failure really, not just a recession, but a full blown depression. And you can see the behaviors that they had. But when I look at my sons going through college now and young people 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 demographics behavior in databases today. And so 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 societal interests, 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 sometimes, to some extent, law, but also because of market interest. This broadens our market score consumers to engage in a successful product.

And again, it's the 50 or the 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, use 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 economic benefits very broadly. So good for that.

PAMELA DIXON: To pick up on Stuart's comments, 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 HIPPA, and folks who are regulated by the common role, 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 Stewart 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 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 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.

TIFFANY GEORGE: Let me just piggyback a little bit on what Pam just said about the richness of the data set. And 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. And does anyone have thoughts on why that is and how it can be addresses?

NICOL TURNER-LEE: Actually, I was going to step right into that puddle there. I think that's an interesting piece because we often think about, in these conversations for those of us who are entrenched in the telecom space, broadband adoption here, data here, broadband enabled applications here. And actually all these verticals cross at some point to give us a rich, robust conversation and story on how all these things interface. And I would say-- and I give a shout out to the Center for Data Innovation. Daniel, I saw him here who published a paper. I didn't get a chance to read, but I got 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 [? Horigan ?] has mentioned, the level of digital readiness to actually go online and engage in a very participative way on the internet for non-commercial value versus commercial value, et cetera. You put all that together. I was thinking about your comments, Stuart. You might begin to see some segmented marketing to some of those folks because you have the other sociologists, the perspective that my online behavior may match what I do offline. And so I may be looking for 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 and 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 and digital divide issue is not going 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, could benefit from the positives of big data. They get left out. Or their results gets skewed because the proportion of people that are participating may not have these other factors, the literacy and the readiness at hand, to equally participate. I think the inclusion piece, Center for Data Innovation's last point calls it the data divide, it still goes back to the data and inclusion divide on how you look at this big picture.

STUART PRATT: I would add, though, one of the approaches our industry has taken, though, whether it's a fraud prevention tool-- by the way, we live very much in the fraud prevention world and the ability to pay world. And really, everything, all that data 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-- and I'm going to use a term that we've used at CDI-- necessary services, so ubiquity. In other words, there's a question of that. In others 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're likely paying for a utility of some sort. It could be very straightforward, water service in this sort, electricity. And then telecom is an example of, again, where you have a fairly ubiquitous set of data. You 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 small to medium enterprise lending. But it ties in with really personal loans as well. It's almost the same thing as coterminus in a lot of places.

But the idea is what data sets are out there-- Colombia, for example, not South Carolina, but Colombia-- uses telecom data widely. This, 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.

CDA 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, however, 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 the CDI sponsored a symposium on this earlier this year. It was hosted by Pew, but it was run by Credit Builders.

I think it's pretty good, intense dialogue. And obviously, 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, the 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 under banked depending on definitions. Under banked have different needs than middle class consumers who are still living in very tight circumstances.

So as you move through societal traunches of consumers, the kind of data that we have allows us to work through that and, again, match up a better offer, we hope. An offer which leads to success on both sides.

TIFFANY GEORGE: 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 during 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.

CYNTHIA DWORK: I think that privacy and fairness are completely unrelated. And I 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 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 getting 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.

PAMELA DIXON: You know I've got to respond to that, right?

CYNTHIA DWORK: Go for it.

PAMELA DIXON: OK. 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 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 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. 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, OK, 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, HIPPA, Common Role, 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 you're 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. HIPPA was right in how it handled that. It handles medical research. For human subject research protection, there is very meaningful robust consent and what's called an IRB process, institutional review board. And so there are examples already in place where we can go.

CYNTHIA DWORK: 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, some kind of a measurement for any particular classification task, searching for a way of measuring how similar or dissimilar are two people for this particular classification task.

PAMELA DIXON: That's right.

CYNTHIA 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 the information. And that's probably very reasonable.

MARK MACCARTHY: So let me jump in here. I think it probably is that this is a very abstract and almost philosophical question if you look at some of Cynthia's work. I was telling her that 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. Emmanuel Kant said that in his hysteria 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 out the big data subset of all data analysis, I think, is the wrong direction to be thinking about. I do think we need the focus not as kind of global solutions to all these problems, but to go back to the specifics.

As Stuart's been saying, 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 the largest 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 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.

NICOL TURNER-LEE: I want to jump in. 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'm a big fan of the FIPS 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 on 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 anymore to actually say, well, I'm going to the internet for this. I'm going for that. 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.

When you give your email address on the internet, 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 is probably a search that you did that brings up a health care provider. You

might have gone and bought red shoes. 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 a hard ecosystem to distinguish between this is why people are going to the internet, for this particular purpose. So I think a general framework, like the FIPS, is actually appropriate to help us figure out, how do you ensure that the input of data, whether it's big or small data, does not impute cultural stereotypes as well as cultural clichés that actually lend itself to predatory behavior and actions on the part of the online space. I think that's so important. We've seen it with segmented marketing where, again, for people of color-- 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 someone has seen something on my Facebook page or I've 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 discriminate against people. The question becomes, do most consumers know that when they participate, particularly from minority consumers who over index in social media when they are on and over index 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 just because they want to be part of the conversation?

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 MMTC. Will that have an impact on someone's ability to get a job or health care or something of social value, not necessarily their ability to stream content, but something of social value that will essentially, when they are applying for a car loan, will give them higher rates. I think that's really important to put in this conversation.

ALESSANDRO ACQUISTI: I wanted to connect what Nicol just said to something Cynthia said. It's something Solon this morning was mentioning. So I'm ready to believe that most of the time, 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. Because the point Nicol was referring to was the experiment we did with the impact that social media information has on the hiring behavior of US employers.

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 attainment and professional achievements for different candidates. However, we had also created social media profiles for this candidate.

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 the disclosure of the religious affiliation in a discriminatory manner. So 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 suggests with potential problem. Sometimes, with more information is 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 legislator who decided to have regulatory protections on certain traits, so that certain traits should not be asked about in interviews or should not be used in the hiring process, and information, normal information technology, which is effectively bypassing the legislation because he's making this new data, these new attributes, perfectly easily available to employers, without employers even needing to ask during an interview.

PAMELA DIXON: 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, one of the difficult things that I was forced to unambiguously ascend to at the conclusion of the researching of the scoring paper is that, really, we cannot control our information flows anymore, our so-called digital exhaust. We really don't have the rights, 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, OK. So who's doing this? Is this just 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, big patterns here. 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 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'll help perhaps your health plan 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's 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 Axiom has one of these, an opt out, 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, categorization is a big deal. And it can really change how your life looks.

CYNTHIA 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. An Pan Datha was one of the authors. But 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.

So I'm sorry I'm not informed in more detail. But I suggest that people look this up.

STUART PRATT: You can see how, in this dialogue, we're beginning to categorize uses as well. 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 kinds of data sets that a CDIA member has are not often and certainly not for risk management purposes big data that is derived from 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-- there's an example of a lender with a closed system of data. And the consumer said, yes, you can use this data. I don't have traditional data sets for you to be able to make that lending decision. So I do think that's occurring.

Also, we haven't talk 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. Unstructured data might be data that's less directly identified with me. Depends on whether you think an IP address is personally identifiable information or not.

CYNTHIA DWORK: Yes, it is.

STUART PRATT: No, it's not. 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 to use a Fair Information Practices template of some sort, and there's many

of them out there. I tend to like Apex better than some others. 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, 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 the 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, verses 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 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 economist described that the German government regulates every step of the process in milk production. So really, forgive the pun, it's 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, 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 meets 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. 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, 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 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 in 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.

TIFFANY GEORGE: That's actually the perfect segue to my next question. 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? What are the guideposts that we should be looking for as we emerge into the future?

MARK MACCARTHY: Let me quickly jump, if I could.

CYNTHIA DWORK: I'll go next.

MARK MACCARTHY: Yes. I think you touched on the two big ones, which are use and harm. This brings us back to the 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 a particular outcome that we want. Sometimes, more information is not so good.

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 were actually performing the music and notice which ones were men and which ones were women. When you put them behind a barrier, so you couldn't tell what the sex was, suddenly it became 50-50. Withdrawing information in that particular situation is 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.

Another example of this, and it goes back to your point, will businesses and others 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 Intello 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, to target your recruitment efforts to try to get at the kind of people who will be qualified for your work and 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 that says, here's how we do it in 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.

PAMELA DIXON: So great question, and I appreciate your comments, Mark, they were very thoughtful. So I want to talk about medical just 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 medical research is intriguing in 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.

ALESSANDRO ACQUISTI: You were asking about what model may work. I have a record as criticizing transparency control mechanisms due to a series of behavioral experiments we have run showing how, for instance, control over personal data or even just the feeling of control over 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 further, a little provocation for the panel. I'll focus on the cost of the data [INAUDIBLE]. What if we start applying the rules of the data industry, the ones we use on consumer data, we apply the same rules from the consumer on the data that firms have about consumers? So imagine a system where every piece of information held by any data holder has to be a companion, 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's 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's presented to us, merely being able to solve in the close future any societal problem, that kind of technology of attaching metadata showing the provenance of personal information is not really that science fiction like.

Otherwise, if you keep adding 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.

TIFFANY 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 to 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 technologies 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.

NICOL TURNER-LEE: This is a very interesting question because the 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 has 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, 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 people to see how often, like me, you don't make it your bed often because you're also reading papers sitting on your couch?

So 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 when we're coming up with a framework, does it balance use versus harm with allowing some flexibility for the collection of data that will help us with the purposes again of efficiency and public good, and the extent to which consumers from the front-- we start talking about this. And not to make this long-winded, but we started talking about this when I was at the 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 Ph.D. or a J.D. just to read the policy after we ran them through the fluency indicator.

The level of what people are engaged in is sometimes not known 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-- we should not have it where we look at consumer protection when a bad actor 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 or credit because of a infraction or harm.

So we have to figure out ways for people to have a lot more knowledge as to, one, the internet is a participatory environment. And in some cases, you'll know why your data is being collected, and sometimes you won't. Two, when I feel that there's some particular harm or some type of compromise in terms of my personal identifiable data in particular, 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 and not necessarily something that's more pertinent to the broader group. So I'll pass it over to you.

STUART PRATT: Thank you. 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-- even if you were looking at a FIPS model, it's not monolithic. I remember working with a GAO group a while ago. They were looking at government uses of data. And they have applied an OECD FIPS 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 FIPS 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 was years ago I remember, 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 was clicking and clicking and clicking trying to get rid of all the damn cookie notices. And before you know it, I was not reading the cookie notices. I was just doing battle with them so I could actually see what was on the screen. So there would be almost a behavioral issue there for consumers.

How do consumers behave? What is your goal? And what's the most effective strategy to get to that goal? So I'd say it's 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.

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, OK, 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.

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

MARK 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, if that's really your focus and you're really pushing that as your major defense against unfairness and privacy invasions, you've got to do it in some cases. Human subject experimentation is not something we want to make decisions for people. But if that's your universal solution, I think you're really doing customers and consumers a disservice.

You're responsiblizing 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 themselves completely.

CYNTHIA 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 classification 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.

PAMELA 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 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 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.

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

TIFFANY 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]