FEDERAL TRADE COMMISSION

COMPETITION AND CONSUMER PROTECTION
IN THE 21ST CENTURY

Tuesday, November 6, 2018
9:00 a.m.

American University
Washington College of Law
4300 Nebraska Avenue, N.W.
Washington, D.C.

1 FEDERAL TRADE COMMISSION
2 I N D E X
3 PAGE:

1 P R O C E E D I N G S
2 MR. GILMAN: Good morning, everyone. My
3 name is Dan Gilman. I am at the FTC's Office of
4 Policy Planning. Just a couple of really short
5 announcements before we get to today's program.
6 First, everyone ought to know that this is a
7 public event not just for your attendance, but it is
8 being webcast. So you are being recorded. There will
9 also be a transcript of today's proceedings taken and
10 then subsequently made available.
11 Number two, some of you may have already
12 gotten question cards on the way in. We have them
13 available throughout the day. People will collect
14 them. Staff will read them all, process them all.
15 Some of them will be passed along to panelists during
16 the day, not necessarily all of them, but we will take
17 them. We are going to try and keep a prompt schedule,
18 if we can.
19 So without spending any more time, I want to
20 introduce -- oh, biographies are available. So we
21 have very, very accomplished people here today. We
22 are not going to recite their accomplishments at you,
23 but the biographies are available.
24 I just want to introduce Professor Jonathan
25 Baker, an antitrust scholar here at American

1 University Washington College of Law for welcoming
2 remarks.
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

1 WELCOME AND INTRODUCTORY REMARKS
2 MR. BAKER: Thank you, Dan. I am delighted
3 to welcome the Federal Trade Commission and the
4 antitrust and consumer protection community to my law
5 school. If you have not been here before, I hope you
6 will stay some time to meet some of our terrific
7 students and admire our wonderful facility where we
8 have now been for about two years.
9 I have served twice at the Federal Trade
10 Commission, once as an attorney advisor to
11 Commissioner Terry Calvani and then later as the
12 Director of the Bureau of Economics when Bob Pitofsky
13 was Chair.
14 When Chairman Simons opened these hearings
15 in September, he said he modeled them on the hearings
16 that Chairman Pitofsky held in 1995 when I was at the
17 Federal Trade Commission. The Pitofsky hearings were
18 prompted in part by two ways the economy had changed
19 since the mid-20th Century. First, markets were
20 increasingly globalized. In the four decades since
21 the end of the Second World War, firms across the
22 developed world, particularly in Europe and Japan, had
23 caught up to their U.S. counterparts. And that
24 created more competition for many domestic firms at
25 home and abroad. And antitrust enforcers were

1 increasingly detecting international cartels.
2 The second change in the economy between the
3 mid-20th Century and 1995 was the growing importance
4 and pace of technological change. You could see that
5 particularly in information technology. This was a
6 decade after Microsoft introduced the Windows
7 Operating System for the IBM PC and we were right at
8 the start of the dot-com boom.
9 The changes in the economy that we saw in
10 1995 are still continuing. International trade has
11 continued to increase as a fraction of GDP, and
12 although the overall rate of productivity growth has
13 probably slowed since 1995, many of what are now the
14 largest internet and information technology firms were
15 just being born then. Amazon was only a year old.
16 Facebook and Google were still to come.
17 The rise of the internet points to new and
18 distinctive challenges for the hearings that the
19 Federal Trade Commission is now conducting,
20 particularly for the ones for this week. The
21 transformation for information technology since 1995,
22 and particularly the growth of online platforms, is at
23 the heart of the novel competition and consumer
24 protection challenges that the FTC must now address.
25 On the consumer protection side, online

1 platforms provide a new locus for fraud and deception
2 and the migration of personal data to online hosts
3 creates new privacy challenges.
4 On the antitrust side, if you credit the
5 recent economic research that suggests that market
6 power has been on the rise for decades, which is what
7 I talked about last month on the opening day of the
8 hearings, then it is natural to ask whether increasing
9 market power is related to the growth of information
10 technology generally and look closely at the conduct
11 of the internet giants, in particular, including the
12 way they develop and use data about their customers
13 and their suppliers.
14 So the issues that the Federal Trade
15 Commission is concerned with this week are at the
16 center of the new challenges for antitrust and
17 consumer protection that are created by the 21st
18 Century economy.
19 On behalf of the American University
20 Washington College of Law, I am delighted to welcome
21 everyone to this important two and a half day
22 conversation.
23 So let me now introduce one of my successors
24 as the Director of the Bureau of Economics, Ginger Jin
25 from the University of Maryland, who will give us an

1 introduction to the economics of big data, privacy,
2 and competition.
3 (Applause.)
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

1 THE ECONOMICS OF BIG DATA, PRIVACY, AND
2 COMPETITION - AN INTRODUCTION
3 MS. JIN: Thank you so much for having me.
4 I appreciate the opportunity to share my thoughts
5 about big data with you.
6 As an economic researcher, I have done some
7 research about markets with asymmetric information,
8 but not data or privacy-specific before I joined the
9 Commission in 2015. However, the precious experience
10 at the Commission has exposed me to a lot of cases in
11 data security and privacy, which pushed me to dig
12 deeper into the market and think hard about the
13 potential benefits and risks related to data
14 collection, data use and data sharing.
15 I remember at that time, when I started this
16 learning process, I felt that I am on a fast-moving
17 train, but I am not sure where it is going. Two years
18 later, even after I had returned to economics, I think
19 the speed of the train has been faster than I thought
20 and the destination is even fuzzier. So as a result,
21 I have a lot of questions in my mind to which a
22 comprehensive and a satisfactory answer is yet to
23 come.
24 I hope hearings like this and before and
25 after this would provide opportunity for everyone to

1 think about this issue, to chime in with their own
2 opinion, and really form a collective wisdom. And
3 this collective wisdom, I believe, would have an
4 impact for our policymakers to make informed
5 decisions.
6 So today, I would just probably organize my
7 thoughts in an economic framework. It probably is not
8 precise to call them thoughts, but just a list of
9 questions and hopefully that will stir conversation in
10 two and a half days of this hearing.
11 So the first question I asked myself is,
12 what is going on in the marketplace? And to begin
13 this question, I want to look at the kind of players
14 in the market. We are all familiar with the role of
15 firms here, but I want to make some comment about
16 consumers, government, and research institutes.
17 So consumers in the data market are not just
18 consuming products and services backed by data. They
19 are also active data providers and data users. How
20 many of you have, say, a smart watch on you sometime
21 during the day? Some of you.
22 So you can see from these kind of devices
23 and online apps that we are constantly providing data
24 to the app. We are also consuming data from that. We
25 want to know the statistics, how many steps we have

1 walked today and how many miles we have run, and so
2 forth. So this is a very active data exchange between
3 consumers and firms. So consumers are not passive
4 sort of consumers of the products generated out of
5 data; they are also actively participating in this
6 process.
7 And to some extent, the Government is
8 similar to consumers. They consume data. They also
9 provide data. However, the Government has the power
10 to make new legislation about this market. They can
11 designate certain law enforcement to enforce the law.
12 So in that sense, the Government is both a player and
13 a referee. So I think that combination probably will
14 make Government's role distinctive from all the other
15 players here.
16 In terms of research institutes, here I want
17 it to be a broad definition, not only economic
18 institute but also, say, think tanks, consumer groups,
19 even industry associations. And those institutes, we
20 are -- as an economic researcher, I can say that I am
21 always hungry for data to make my research more
22 insightful. But on the other hand, we also want those
23 research institutes to be kind of a third party to
24 describe the marketplace to us from an objective point
25 of view. So I think that role probably individual

1 consumers cannot play, but will be very important in
2 this marketplace.
3 In terms of exactly what is going on, I hope
4 this hearing and other hearings would shed more light
5 on who generates most data; who uses which data for
6 what purpose; where and how does data stay, flow
7 and evolve; and how does technology reshape data
8 and data use; who benefits, who loses from certain
9 data practices; and what is the aggregate consequence
10 of data use in the short run and in the long run;
11 and what is known and what is not known, to whom and
12 when.
13 I really think those questions have to be
14 addressed by probably a multidisciplinary approach,
15 not only from the Commission's own research report,
16 which has been done in 2014 and 2016 about data, but
17 also from, say, computer scientists, economists, law
18 professors or even psychologists, to really help us
19 understand how each player works in this space. I
20 would encourage all the think tanks and organizations
21 to contribute to this, as well. Of course, firms
22 should give us probably a more intimate view of
23 exactly what they have been using the data and what
24 thoughts they have had when they decide the policies
25 about the data use. So I hope this afternoon's

1 session about the business of big data would really
2 give us more insights on this.
3 So suppose we sort of understand how the
4 market works, probably we should ask, is there
5 something wrong and what goes wrong? And as an
6 economist, I often try to think of that question as
7 where does the market fail? We cannot just say this
8 is an issue and then jump directly into intervention.
9 We probably have to ask to what extent that the market
10 is able to address that question, okay, and then where
11 the market is not able to address that question.
12 So following that line, I am thinking about
13 the textbook examples of market failures and there are
14 typically four of them. The first one is well known,
15 market power. There is a long history of antitrust
16 talking about this in monopoly and oligopoly, market
17 structure. The second one is information asymmetry.
18 The third one is externality. The fourth one is
19 bounded rationality.
20 And I want to push the audience to think
21 exactly whether and how does big data contribute to
22 these market failures, okay? I want to be a little
23 specific. For example, if you think about potential
24 market failure from market power, does data constitute
25 barrier to entry? Does data facilitate conclusion

1 between oligopolic firms? Does data facilitate
2 anticompetitive contracting? Does data facilitate
3 perfect price discrimination? And on the other side,
4 data could also generate merger efficiency or contract
5 efficiency.
6 Based on my experience, I think the
7 potential anticompetitive practice related to data is
8 more often a theoretical possibility than a widespread
9 practice in the real world. I am happy to be
10 corrected by maybe tomorrow's panel discussion on
11 this, and if there are more evidence towards
12 anticompetitive direction, I will be really happy to
13 be corrected.
14 So if we identify some contribution of big
15 data to the anticompetitive problem I listed here, I
16 think that still has to be translated into what is the
17 overall impact of that practice on consumer welfare,
18 both short run and long run. That is sort of where
19 the real and tangible harm should be associated with
20 big data before we take antitrust action towards that.
21 Okay. The second one is information
22 asymmetry. I know not all of you have economic
23 training here. A very textbook example about
24 information asymmetry is prescription drugs. That is,
25 we, as consumers, we do not know exactly what is in

1 that particular pill. The firms could probably do
2 some advertising telling us that, okay, we really have
3 a cancer cure in that tablet. However, even after we
4 consume it, we still cannot tell whether it has really
5 cured our cancer because there are so many other
6 things going on. So this is a very typical
7 information asymmetric problem because the firms know
8 more about the product than individual consumers.
9 If we sort of borrow that kind of mind set
10 into the data-related issues, then I would say the
11 information asymmetry associated with data is probably
12 even more complicated than prescription drugs in the
13 sense that we not only have information asymmetry
14 before the focal transaction, consumers do not know
15 how they are going to use that data for the particular
16 transaction, for example. But, also, a lot of
17 asymmetry would arise after that focal transaction.
18 We do not know how the firm is going to store the
19 data, to what extent they are going to change the
20 content and format of the data, and to what extent
21 they are going to sort of link the data with something
22 else, okay?
23 This is not only just the information set of
24 consumers at the point of focal transaction or after
25 the focal transaction, but, also, sort of what is the

1 information set of firms as time goes on, right? They
2 may not know exactly what they are going to do with
3 the data, but they will have some say in how they are
4 going to use the data later on. And that question
5 also relates to affiliates or even nonaffiliates of
6 the firm if they are going to share the data with the
7 firm.
8 And I would also add black-market players
9 like hackers and the public here because we know in
10 incidents like data breach and other things, that --
11 maybe this is an unintended data use, but it turns out
12 to be a potential data use in reality.
13 So coming back to this core question, what
14 is the harm to consumer welfare from the information
15 asymmetry problem of data and where does it show up
16 and how much is it? Can we really quantify it?
17 So the third market failure, the potential
18 market failure, is externality. What is the typical
19 example of externality? Let's say air pollution,
20 right? We could have a lot of firms producing harmful
21 gas into the air. We, as, say, the general public or
22 the consumer of air, we sort of probably can tell the
23 air does not smell right and we can do some lab tests
24 showing that there are some harmful components in the
25 air, but we do not know exactly which firm contributes

1 to that air pollution.
2 And this negative externality is not taken
3 into account by the firms in their market practice
4 which generates this negative externality problem. If
5 we bring that mind set to the data issue, there could
6 be questions like, what data practice would generate
7 what spillover? And we know that according to the
8 Bureau of Justice statistics, about 7 percent of
9 American people above the age of 16 is a victim of
10 identity theft, and a lot of identity theft are
11 related to data issues.
12 However, even if I am a victim of identity
13 theft, I do not know exactly which of the hundreds of
14 firms I interacted with in my past will sort of really
15 contribute to this event of identity theft. In that
16 sense, it is kind of a similar problem of negative
17 externality as the air pollution I just talked about.
18 Okay? So that is just negative externality.
19 There could also be positive externality in
20 the sense that we know if a lot of data sets pulled
21 together would really help, say, the census or
22 researchers using the census being able to generate
23 research grade outcomes. However, each firm may not
24 have the full incentive to share that data because
25 they are not going to get all the returns from that

1 data use. So in that sense, we could even have
2 positive spillovers which generate an under-incentive
3 to collect and share data.
4 So I want this hearing -- I am hopeful that
5 this hearing will talk about the externality issues in
6 data and to what extent the parties that generate that
7 spillover have the incentive to internalize that
8 spillover and how does that spillover affect consumer
9 welfare.
10 So the last potential market failure is
11 bounded rationality. We know a lot of us have been
12 sophisticated, but we are not as sophisticated as the
13 machine could be or as a rational agent in an economic
14 model would assume. So we always have some level of
15 sort of standard rationality or you can say the
16 rational choice of not paying attention. And this
17 could happen in this area.
18 And we know, thanks to researchers like
19 Laurie Kernoff (phonetic) that -- we know ten years
20 ago that very few people actually read privacy policy.
21 However, we still have that as one of the main
22 building blocks for today's data space. So exactly
23 how consumers, how individuals deal with this kind of
24 information presented in front of them when they have
25 very limited attention, but a lot of information to

1 digest. Okay?
2 On the other hand, firms probably are hungry
3 for data and they have more resources to deal with the
4 data and they can employ or even invent technology to
5 process data. So in that sense, my view is the
6 asymmetric information between the consumers and the
7 firms have been magnified by this advance. On one
8 hand, the consumers are driven by inattention, they
9 want quick and straightforward solutions. On the
10 other hand, the firms are really churning up a lot of
11 resources and technology to try to digest as much
12 information as possible.
13 So that brings a question of who has more
14 bounded rationality in this marketplace? Who suffers
15 from bounded rationality and whether some parties
16 would have incentive to exploit other people's bounded
17 rationality. And, again, I want this to sort of boil
18 down to exactly how does this bounded rationality
19 affect consumer welfare.
20 Okay. So that is kind of market failures
21 from the economics point of view. And suppose we
22 identify one or more market failures in this area,
23 then we could talk about a bunch of potential
24 solutions. Here, I am putting kind of a spectrum from
25 free market to having prescriptive regulation from the

1 Government. Okay? So in the middle, we could have
2 industry self-regulation, some guidance to the
3 industry firms and somehow there is a mechanism for
4 firms to conform with that, or we can sort of
5 strengthen that by more external monitoring, like the
6 consumer education effort, as well as societal
7 monitoring, and all these probably not involve
8 government.
9 If we could push it a little bit further, we
10 could have government involved in ex-post enforcement
11 and that is kind of like, say, nutrition supplements,
12 right? Okay, you can put the nutrition supplements in
13 the market without going through the FDA and clinical
14 trial. But if something goes wrong with that, then
15 law enforcement effort would come in and to try to
16 correct that. So that is probably less aggressive
17 than the FDA approach, say, in food labeling or drug
18 clinical trials.
19 And that brings me to the ex-ante
20 regulation, that we could have heavy-handed regulation
21 like define exactly what you can say, what you cannot
22 say, we are going to find a way to confirm that what
23 you said is correct. We can sort of inspect you
24 saying you have to do A, B, C before you produce a
25 product because we believe A, B, C is kind of good in

1 ensuring the quality in the final product or we can
2 even impose a minimum quality standard on the final
3 product you eventually produce, like a clinical trial
4 to make sure that a drug is safe and effective in
5 addressing certain diseases.
6 We can combine both the ex-ante regulation
7 and ex-post enforcement and sort of having this in a
8 dynamic sense that we can revise our legislation given
9 the new questions coming out and so forth. So I want
10 you to have this spectrum in your mind when you think
11 about what is the potential solution and what is the
12 tradeoff of each solution.
13 So now, suppose we sort of agreed on which
14 solution we are going to get, and then the question is
15 exactly how we get to the ideal effect of that
16 solution. I have heard people talking about using
17 existing rules, such as competition law and consumer
18 protection law. And I guess the immediate question
19 is, how do they fit in this overall framework I just
20 discussed about market failures and the potential
21 solutions?
22 And the second question is, what is the
23 relationship between the two poles, okay? They could
24 be sort of -- let’s say on your left-hand side, I put
25 it as a leverage, like the two could be conflicting

1 with each other. Let me give you an example. So
2 antitrust may concern about data not available to a
3 potential entrant into the market and, therefore, push
4 for data access, data portability, and data
5 standardization. However, the consumer protection
6 part may worry about that there might be some
7 unintended use of the data and, therefore, the
8 consumer should have a right to restrict how their
9 data should be used. And that could generate an
10 effect that actually reduces the potential entrant's
11 access to the data and the data portability.
12 So in that sense, these two may be just sort
13 of contradicting with each other. Is that the world
14 we live in that we have to find the balance point
15 between the two or maybe we sort of need the two gears
16 to work together?
17 Let me give you another example. Say we
18 have a lot of data policy, they are very long, legal
19 language and hard to understand. If there is no sort
20 of consumer protection enforcement on how clear this
21 policy must be -- and firms may find that the more
22 obscure the language, the better I can get data and
23 really benefit from it, and then promoting
24 competition, actually would push firms to compete in
25 that particular dimension, which means the data

1 available to consumers -- the data policy available to
2 consumers become more and more obscure. So we could
3 talk about like competition in the wrong dimension.
4 So in that sense, we want the two gears to
5 somehow work together in a complementary way. So I
6 hope the hearing would sort of promote a discussion on
7 exactly what is the relationship between these two
8 existing tours.
9 Okay. So there are a lot of questions on
10 exactly how to exactly carry out the solution. I
11 would just list some questions here for the base of
12 discussion. For example, should we aim for the
13 legislation to be very comprehensive and detailed or
14 shall we leave the detail to the regulatory and
15 enforcing agencies? There are arguments in both
16 ways.
17 Who should be this regulatory or enforcement
18 agency? Should that be one or should that be multiple
19 agencies? Should that be sort of at the federal level
20 for everything or should that be at both federal and
21 the state level or just the state level? Should we do
22 this industry-specific or should we cover all
23 industries? And there are questions like the degree
24 of enforcement and regulatory freedom, the resources
25 and expertise available to this or these enforcement

1 agencies.
2 I want to make the extra point here that
3 whatever the agency that the Congress have determined
4 to give power to, assuming that we sort of agree that
5 it is necessary to have such an agency to do their
6 enforcement and regulatory function, I think we should
7 think hard about how do we to limit the agency's power
8 in terms of should we define who this agency should
9 report to, how transparent their practice should be,
10 and how can we make sure that this agency's action is
11 accountable. If they do something over the defined
12 area, how can we correct it and how can we bring
13 external forces to really spot and correct those kind
14 of wrongdoings?
15 So in that sense, I hope other parties will
16 be able to contribute to that solution, even after we
17 have decided exactly how to carry out that solution.
18 And given how fast technology is moving in this area,
19 I think it is really, really important for all the
20 parties I listed here to continue contributing to that
21 solution on an ongoing basis.
22 I only have two minutes left so let me make
23 the final comment about international complications.
24 Every country is doing this slightly differently. I
25 think, to me, there are sort of three models at least

1 coming out of this heterogeneity. One is the European
2 model, that they have a comprehensive framework
3 covering all countries in the EU, which is GDPR, and
4 they have DG-comp in the antitrust agency for the EU.
5 But they also have country-specific enforcement,
6 especially for GDPR. Okay? So that is one model.
7 Another model is sort of the U.S. status
8 quo. We have a patchwork of federal, state, and
9 industry-specific enforcement and they generate some
10 heterogeneity even within the U.S.
11 And then the third model is the China model.
12 They have nationwide laws in 2017, I think. We do not
13 know exactly how they are going to enforce that yet.
14 But we also know that big data could be an input for
15 government censorship and surveillance there.
16 So I am not saying that I have a good idea
17 of which model of these three is good or is better
18 than others, but I think it is really important to
19 discuss the pros and cons of these approaches. This
20 is not only because companies are global and they have
21 trouble conforming with all kinds of different
22 regimes, but also because -- I think this is more
23 important -- but also because data, ideas, talents,
24 and the money flow globally. Okay?
25 So that means if in one corner of the world

1 they have very prescriptive regulation, maybe the
2 money and talent and idea would go somewhere else,
3 okay? And what is the implication of that for the
4 whole economy in terms of consumer welfare, as well as
5 the future innovation and support of the economy. I
6 think that is a very big question. So I am going to
7 stop here.
8 Thank you very much.
9 (Applause.)
10 MR. GILMAN: Thanks very much, Ginger. We
11 have a break scheduled now. I would just ask you are
12 getting out a little bit early because we started a
13 little bit early. I would ask people to be in their
14 seats promptly at 10:00, so we can start again on
15 time. Thanks very much.
16
17
18
19
20
21
22
23
24
25

1 THE ECONOMICS OF BIG DATA AND PERSONAL INFORMATION
2 MR. SANDFORD: Okay. Good morning to those
3 in the room and those watching on the webcast. This
4 is our panel on the economics of big data and privacy.
5 We have five panelists here to share their views on
6 how markets involving big data and privacy function.
7 We have Alessandro Acquisti from Carnegie
8 Mellon University. We have Omri Ben-Shahar from the
9 University of Chicago Law School. We have Liad Wagman
10 from the IIT Stuart School of Business in Chicago. We
11 have Florian Zettelmeyer from the Kellogg School of
12 Management at Northwestern University. And we have
13 already heard from Ginger Jin, who is from the
14 University of Maryland.
15 My name is Jeremy Sandford. I am an
16 economist at the Federal Trade Commission. I work in
17 antitrust, and for the most part, my colleagues in
18 consumer protection at the agency are those that deal
19 with big data and privacy issues. So, hopefully, this
20 mismatch is a feature and not a bug.
21 The reason we have an antitrust person
22 moderating this panel is, well, there have been calls
23 for increased antitrust enforcement of big data and
24 privacy issues. So, for example, Joe Stiglitz,
25 speaking at an earlier hearing, shared his view that

1 big data and privacy represent one of the biggest
2 challenges to our society and to competition law. So
3 we kind of want to get at the question of should we be
4 doing something different with respect to antitrust
5 when we have, say, a merger or single-firm conduct
6 that involves big data or privacy.
7 My focus on competition is not a constraint
8 on the panel or their opening statements. You all can
9 talk about whatever you want and we are going to hear
10 from our panel on kind of their views on how these
11 markets work. And then I am going to ask questions
12 that are going to kind of get at are there competition
13 implications for big data and privacy markets that we
14 may not be taking into account with the way we do
15 things now.
16 Okay. So we are going to proceed as
17 follows. We have already heard from Ginger, so she is
18 not going to speak again. But each of the four
19 remaining panelists will have up to ten minutes for
20 opening remarks and then we will have a Q&A session
21 where I will ask questions and the panel will answer.
22 If you are in the room here at American
23 University and you would like to ask a question of the
24 panel, we will have people going up and down the
25 aisles with note cards. You can flag one of them

1 down, get the note card, write your question on the
2 note card, and someone will bring it up to me, and I
3 will see what I can do of asking those questions.
4 So the order of speakers will be
5 alphabetical. So we will have Alessandro, Omri,
6 Florian -- sorry. Alessandro, Omri, Liad and then
7 Florian.
8 MR. ACQUISTI: So good morning and thank you
9 so much for the invitation. And, more importantly,
10 thank you to the FTC and American University for
11 creating this forum. The quality and diversity of the
12 speakers is -- should I push something?
13 Thank you so much. So I guess you heard my
14 thanks. And I was adding that the quality and the
15 diversity of the speakers is exactly what we need to
16 bring nuance and some degree of clarity to a complex
17 topic.
18 And in my remarks, I will focus on two
19 different areas. First, I will go broad and propose
20 some personal framings, some ways to frame the debate
21 over big data and privacy. And I will focus in doing
22 so on two apparent issues, yet common misconceptions,
23 which we, as scholars, are aware of, not often they
24 are properly understood in the public debate over
25 privacy.

1 Second and next, I will go narrower and I
2 will present some ongoing, yet unpublished, work we
3 are doing on the topic of the allocation of value
4 created by the data economy. Okay?
5 So starting from the framing of the
6 misconceptions, the first misconception is that
7 privacy and analytics are antithetical. You can have
8 one or the other, but not both. You find echoes of
9 that stance already back in the days in the writings
10 of scholars whom I actually greatly admire and respect
11 because they were the first scholars to bring
12 economics to the field of privacy, Chicago School
13 scholars such as Posner and Stigler, who conceive of
14 privacy as effectively the concealment of information,
15 the blockage of information flows.
16 Now, we know from the case of work on
17 privacy that a much more nuanced, and I would say,
18 precise view of privacy is in terms of management of
19 information flows, not blockage. It is -- sharing a
20 secret with a friend or posting some information on
21 social media and choosing the visibility setting for
22 the post are sharing behaviors, which are also privacy
23 behaviors. They are privacy behaviors because they
24 encapsulate the ability to manage the boundary between
25 the self and the others, which is far from the notion

1 of privacy as a blockage of data.
2 Why is this important? It is important
3 because once you realize there is more -- in yourself
4 there is more than one view of privacy as management
5 of this boundary between privacy -- between private
6 and public, then you also realize that it is, in fact,
7 possible to have simultaneous privacy in analytics to
8 protect certain types of data and share certain types
9 of data.
10 We can do so through truly an actionable,
11 informed consent, something that I do not believe is
12 very common nowadays in the privacy landscape. We can
13 do so through smart regulation. We can do so through
14 privacy-announcing technologies. The best of these
15 technologies do not block data; rather, they try to
16 modulate what data is protected, what data is shared
17 in the interest of increasing welfare of different
18 stakeholders.
19 The second and a related misconception is
20 that the relationship between data protection and
21 generation of economic value is a monotonic,
22 specifically data protection is always welfare-
23 decreasing and data collection is welfare-increasing.
24 In reality, both in theory papers and empirical ones,
25 we have a much more nuanced view and we realize that

1 the economic impact is very much context-dependent.
2 For instance, healthcare privacy regulation,
3 if done improperly, could slow down technological
4 innovation in healthcare -- Amalia Miller and
5 Catherine Tucker have important papers in this area --
6 but if done properly can actually increase innovation,
7 which is something that we found and published in
8 Management Science with Idris Adjerid and Rahul
9 Telang. Social media can lead to better matching in
10 labor markets, but can also lead to more
11 discrimination in labor markets. So it is always
12 context-dependent and we should be very, very cautious
13 about taking a one-size-fits-all when we think about
14 the relationship between data and economic value.
15 I can offer you two further examples of this
16 from scholars who certainly cannot be accused of being
17 against efficiency and against data. The first
18 example is again from scholars I admire from the
19 Chicago School, in particular Posner again, who
20 noticed already in 1981 that privacy is
21 redistributive. The point he was making was that data
22 protection creates economic winners and losers. Now,
23 I believe he is right, but it also turns out that the
24 lack of data protection also creates economic winners
25 and losers. You just cannot avoid this.

1 And the second example, which is related to
2 the first, is from Hal Varian, who in 1996 pointed out
3 how consumers may rationally want marketers to know
4 their preference so they get offers which are of
5 interest to them. But they also may rationally not
6 want marketers to know their willingness to pay in
7 order to avoid being price-discriminated. The first
8 desire is welfare-increasing for the consumer; the
9 second is to avoid a situation which is welfare-
10 decreasing.
11 So the lesson here is to be watchful of
12 arguments, such as data protection is monotonically
13 increasing or decreasing value. The reality is much
14 more nuanced and context-dependent, which brings me to
15 the second part of the talk, where I present some
16 ongoing results from studies we have been doing trying
17 to disentangle these nuances.
18 I will focus in particular on targeted
19 advertising. The reason is that targeted advertising
20 is afflicted by what I was referring to earlier at the
21 beginning of my talk, some of the misconceptions in
22 the public discourse over big data and privacy. There
23 is a sort of magical thinking happening when it comes
24 to targeted advertising, which is reflected in the
25 following words. I am going to cite some words. I am

1 not -- in the privacy spirit of the panel, I am not
2 going to cite the person who wrote them because I do
3 not want to make this an attack on the person. It is
4 a critical argument.
5 Targeted advertising is not only good for
6 consumers. It is a rare win for anyone. It ensures
7 that ad placements display content that you may be
8 interested in rather than ads that are irrelevant and
9 uninteresting. Advertisers achieve a greater chance
10 of selling the product. Publishers also win because
11 behavior targeting increases the value of the ad
12 placement. So basically, everyone benefits from
13 this.
14 Now, at first glance, this seems plausible.
15 The problem is that upon further inspection, you
16 realize that there is very little empirical validation
17 in all these claims. I am trying to choose my words
18 carefully. I say there is very little empirical
19 validation. I did not say that there is a disproof.
20 What I am saying is that we actually do not know very
21 well to what extent these claims are true and false.
22 And this is a pretty big problem because so many of
23 these claims are actually accepted unequivocally and
24 they are quite influential in the public debate over
25 privacy.

1 Why am I claiming that we actually do not
2 know whether these statements are correct? Two
3 reasons. The first reason is that, for all the focus
4 on transparency, the data economy is remarkably an
5 opaque economic black box. For the outsiders -- and
6 outsiders could be maybe the merchant buying online
7 ads or the publishers showing on their websites the
8 ads -- it is very difficult to know what happens
9 inside a black box of the different ad exchanges.
10 And we have evidence of this from lawsuits
11 and scandals, which have arisen repeatedly in the last
12 few years. The Guardian finding out that Rubicon, an
13 advertising firm, retained substantial undisclosed
14 funds, in addition to the fixed percentage fees. We
15 found -- another example of that with Index Exchange,
16 which was using bid caching and gaming auctions for 50
17 percent of impressions. We find evidence of that in
18 Facebook hiding inflated video ad metrics about ad
19 watching for over a year and these metrics of ad
20 watching were inflated up to 900 percent. So that is
21 worrisome.
22 The second reason why I claim that we have
23 little validation for one side or the other of the
24 argument is that much of the seminal groundbreaking
25 and high-quality work in this area on targeted

1 advertising from academia focuses, and necessarily so,
2 on very narrow goals, such as what happens if we use
3 targeted advertising rather than untargeted
4 advertising? Are consumers going to click the ads
5 more? And are the merchants going to see a higher
6 commercial rate? And the answer is typically yes and
7 yes. And this is an important, valuable answer.
8 What that answer misses, however, is the
9 broader picture. What happens in the overall
10 ecosystem? What happens to consumers who do not see
11 those ads or if they see them, what happens if they
12 end up buying something? What would happen, what is
13 the counterfactual if the agency in the ad would have
14 bought a similar good or a higher-priced good or a
15 good with a lesser price, higher quality, lower
16 quality? What happens to the merchants when they
17 start getting engaged in a prisoner’s dilemma style
18 dynamics where they have to use targeted advertising
19 because otherwise their competitors will be poaching
20 consumers away from them precisely using target
21 advertising?
22 So I am referring to more general economic
23 equilibrium kind of analysis. And this is what we
24 will be trying to do recently as well for the past
25 couple years in my research team.

1 I will end by mentioning very briefly the
2 research we have been doing. One year ago, at
3 PrivacyCon, we presented some critical work suggesting
4 that when you account for the different type of data
5 that ad exchanges can use and share with merchants,
6 you will have varied welfare implications for
7 different stakeholders, consumers, merchants and other
8 exchanges.
9 Since then, we have been doing empirical
10 work and I will give very brief examples of these
11 studies. In one study, we have done a lab experiment
12 seeing how consumers react in the presence or absence
13 of ads when they search and try to buy products
14 online. We found that actually there was no
15 difference in amount spent and the satisfaction with
16 the products purchased in the presence or absence of
17 ads.
18 In the second study, we have been gathering
19 data about the prices for goods in organic search
20 results and sponsored search results. We found that
21 prices for goods are, on average, slightly lower in
22 sponsored search results. However, the lowest prices
23 are more likely to be found in organic search results
24 rather than in sponsored search results, so for the
25 ads.

1 And, finally, we have been doing work with a
2 large American publisher from which we got millions of
3 transactions related to the ads they show on their
4 website. We were trying to see how much more revenues
5 they get from ads which are behaviorally targeted
6 versus those that are not. We can do that because we
7 can see whether the visitor added a cookie or not. In
8 the absence of the cookie, it is not possible to
9 target the ad.
10 What we found is that, yes, advertising with
11 cookies, so targeted advertising, did increase
12 revenues but by a tiny amount, 4 percent. In absolute
13 terms, the increasing revenues were $0.0008 per
14 advertisement. Simultaneously, we were running a
15 study as merchants buy ads with different degree of
16 targeting, and we found that for the merchants and
17 buying targeted ads over untargeted ads can be 500 --
18 sorry, 500 percent times as expensive.
19 So although these -- we have to be careful
20 in comparing the numbers -- nevertheless, I leave with
21 the rhetorical question for all of you to consider,
22 which is how is it possible that for merchants, the
23 cost of targeting ads is so much higher whereas for
24 publishers, the return increased revenues for targeted
25 ads is just 4 percent.

1 Thank you.
2 MR. SANDFORD: Thank you, Alessandro.
3 (Applause.)
4 MR. SANDFORD: We will now hear from Omri
5 Ben-Shahar.
6 MR. BEN-SHAHAR: It is always fun and a
7 challenge -- it is not always -- they did not have
8 many opportunities, but it is fun and a challenge to
9 go after my world's all-time favorite privacy
10 researcher, Alessandro, and it sounds fascinating. I
11 should give you my time to tell more about what you
12 are finding because this is really interesting.
13 I guess, first, I want to apologize. I will
14 speak and participate in the panel, but about half an
15 hour before it ends, I have to run to the airport. I
16 have a 3:30 class that hosts a speaker in Chicago that
17 I cannot miss. But thank you for inviting me to take
18 part in this.
19 I am not really a privacy expert. I guess I
20 was invited because I circulated this summer a working
21 paper titled "Data Pollution." I thought I was the
22 only person who thought about it until I heard Ginger
23 also discuss the idea of pollution as a metaphor to
24 thinking about what is the problem that we want to
25 address before we identify how we address it. And so

1 I will briefly discuss what my thinking is in this
2 context.
3 So data policy is focused on privacy, on
4 harms, potential harms, potential injuries, potential
5 reduction in well-being for the people whose data is
6 being taken, used, shared, lost, and so on. And I
7 suggest that there is an additional perspective that
8 can be used to understand the discomfort that people
9 report that they have with the data economy, and that
10 is that the data that is being collected and used,
11 that databases affect others not in these databases,
12 affect an environment, affect an ecology, affect
13 individuals who are not part necessarily to that data.
14 So there is potential negative externality.
15 I would also want to save a minute to talk
16 and to think about externality as a problem not just
17 of negative but also positive. Data has immense
18 positive externalities.
19 What got me to think about this, for a
20 while, I have been kind of -- my area is consumer
21 protection, consumer transactions, consumer contract
22 law. But I have been kind of trying to chime in on
23 debates on privacy, data privacy. I have found that
24 the thing that drives most of what -- of my thinking
25 is what is known as the privacy puzzle, that there are

1 -- privacy experts and advocates really want to do
2 something about a phenomenon that most users seem to
3 be indifferent about.
4 They might say in surveys that they want
5 data to be regulated and that there is a problem and
6 -- but they behave as if there is not, and personally,
7 I was very uncomfortable in the aftermath of the
8 Cambridge Analytica and those in the Facebook fiasco.
9 And I asked myself, what is going on? Why is
10 everybody talking here about privacy when the problem
11 is something bigger than the harm to the individuals
12 whose data was used and circulated to make political
13 lies more effective, that the harms were greater than
14 the harm to these individuals.
15 Namely, there is a problem of -- I thought
16 of it then of pollution, of an entire environment,
17 ecology, being harmed by the practice. Then I started
18 looking and finding many other examples in which this
19 is the -- a year ago there was the Strafa fitness app
20 case, in which it turns out that people share where
21 they run and swim and jog and bike, but you can see
22 where there are clusters of users including American
23 troops outside Niger or in Afghanistan or places like
24 this, not good for national security or for the group
25 as a whole. But, again, it is a problem of public

1 good, not of a private good that is affected.
2 A lot of the -- I also thought that a lot of
3 the data security breaches, Equifax to name one,
4 represent not so much a private harm, but a public
5 good harm. Most people whose data was lost will not
6 be harmed. Those that will be harmed will have -- a
7 lot of it is insured in one way or another. There is
8 -- I do not want to diminish or miscount the important
9 insecurity that is being sensed, but there is an
10 insecurity that is shared by everyone. It is kind of
11 a public -- it is a sense of a degraded environment
12 again.
13 So if the problem is not a problem of
14 externality, you want to think about it in the way
15 that we have been trained to think about
16 externalities, and there is a great model. Data is
17 just the new -- now, this is a cliche by now, but it
18 is just a new fuel. So let’s think about the carbon
19 fuel of the 20th Century and how in the 1960s and ‘70s
20 and ‘80s, regulation began to take over private law as
21 the method to curb the problem of externalities from
22 carbon pollution. We realize that tort suits are
23 failing.
24 And we are realizing now, if you look
25 around, and I can -- you know, many lawyers can attest

1 to that, tort suits in the context of data harms are
2 largely failing, because it is hard to prove causation
3 when Equifax loses your data, how do you know that you
4 are harmed, that your identity theft is related to
5 that and not to something else? The latent effect of
6 the harm and the slow gestation period, exactly the
7 same doctrinal reasons that we had the failure of tort
8 law in the pollution context is failing now.
9 Contracts, of course, are not going to solve the
10 problem of an externality. People are not going to
11 contract for low-emitting products whether they emit
12 carbon or data pollution.
13 So it is -- part of what I did in my study
14 is look at the case law in the era that led to the
15 emergence of environmental law and the EPA, the
16 private law failure that led to that emergence. And I
17 see fantastic parallels from the analytical point or
18 the conceptual point of view to the situation of
19 private law today in an attempt for lawsuits to take
20 -- to regulate the data economy.
21 So if private law fails, maybe for the same
22 reason that it failed in the carbon pollution context,
23 maybe the regulatory approach to environmental -- to
24 industrial pollution should enlighten us into thinking
25 about how to deal with data pollution with the

1 external harms that data produces, and this is maybe a
2 little bit similar to how Ginger previously, at the
3 end of her slide, presented it, but I want to say a
4 few things that were not there, although you probably
5 could foresee them.
6 Environmental law uses three basic
7 regulatory tools, command and control, quantity
8 restrictions. You can only pollute so much. You can
9 only produce so much. Carbon tax, Pigouvian tax, and
10 liability. Now, the GDPR is a type of first -- the
11 first version. Right? Data minimization, data
12 localization, what data you can collect and what you
13 cannot do, this is probably the right way to deal with
14 some of the problems, the problem -- the concern is,
15 of course, that in this area is that it is hard to
16 foresee the problems that will arise and to restrict
17 data only to places where it is harmful and not to
18 also wash out all the potential -- the good effects of
19 data, the immensely good effects of data.
20 So it is a -- you know, while obviously that
21 is part of the solution, it is a very risky solution.
22 It has high -- some benefits, but could also have high
23 cost on innovation. So I tried to focus instead on
24 solutions that were not yet developed in the privacy
25 context to think about the data public harm context.

1 So one is data text. Now I know it sounds a
2 little bit crazy. I am just kind of throwing a
3 benchmark idea. What if we could -- if people use
4 data to pay instead of cash, to pay for the services,
5 for search, for social media? Cash is costly. You
6 use it to pay. You cannot buy other private goods.
7 Data, you can keep paying with it and create negative
8 externalities, share the data about your friends,
9 share -- let Gmail collect the data about messages you
10 got from others who are not Gmail users, things like
11 that that affect others. People seem to be largely
12 oblivious to using that and they should not be.
13 So conceptually -- it is very hard to
14 implement, but conceptually, that problem could be
15 solved by a data text, not a data text that the
16 collectors necessarily pay but that the users that use
17 data as currency have to pay. Now, it really does not
18 matter from an economics point of view who pays for
19 the seller or the buyer. The transaction has to be
20 taxed.
21 This is not a transfer of payment from one
22 site to another to change the distribution of wealth.
23 It is to solve the problem of negative externality.
24 So that is one idea that I put out in the paper, that
25 I set out in the paper, examine a lot of

1 implementation issues. And I do not propose it. I am
2 just saying that this is one way to think about the
3 social cost of data.
4 Another aspect is to think about liability.
5 The third form of regulatory -- third regulatory
6 technique is liability. And here I am thinking about
7 -- mostly about nonintentional omission of data,
8 namely data loss, data security breaches. It is very
9 hard to hold these companies liable for -- it for -- I
10 said in private law, but we do think that there is and
11 I think the FTC -- I have seen previous FTC reports
12 about the estimated social cost of these data
13 emissions so why not use something that has been
14 developed in the pollution context, and that is
15 proportional liability.
16 You do not pay to this victim her actual
17 harm, but when the activity that creates the potential
18 loss, the externality occurs, there should be payment
19 out by the tortfeasor, by the injurer -- it does
20 not matter who it goes to, to the FTC, to the
21 Government -- a fine that represents the expected
22 harm.
23 So here, too, we have to come up with a
24 measure of what is the average cost to a user, to a
25 consumer whose information Equifax lost. It could be

1 a few hundred dollars. It could be less. It could be
2 $10. But there are 143 million of them. So something
3 has to be borne by Equifax, which currently is very
4 hard to do in private law. So I talked about data tax
5 and proportional liability.
6 I will end by saying that I think that this
7 framework helps resolve one of the kind of nagging
8 problems in thinking about data policy and that is the
9 well-known privacy puzzle. Why do people say that
10 they care about data security and data privacy and
11 behave as if they do not? Well, my suggestion is that
12 they are saying that they care about something about
13 the ecology as a whole, about the environment. People
14 can be environmentalists and still fly in and out from
15 Chicago to D.C. for every panel and use a lot of
16 carbon.
17 (Laughter.)
18 MR. BEN-SHAHAR: The private behavior does
19 not necessarily tell us about the extent in which we
20 all believe that there is a public pollution problem
21 to be dealt with. Thank you.
22 (Applause.)
23 MR. SANDFORD: Thank you, Omri.
24 We will now hear from Liad Wagman.
25 MR. WAGMAN: Thanks for having me. So I

1 want to talk a little bit about costs and Omri talked
2 about the costs of data. I want to talk about the
3 costs of privacy.
4 And I started studying privacy from a
5 modeler's perspective. I modeled consumer surplus as
6 a function of, say, privacy regulation or the cost of
7 privacy. So imagine you could have the strictest
8 regime where everybody has privacy. Everybody is
9 anonymous, say, in front of sellers. Or you could
10 have something in the middle where everybody can
11 choose to become anonymous. Or you could have
12 something on the other far end where everybody is
13 known. Okay?
14 And the result of this kind of modeling
15 showed that consumer surplus is not necessarily
16 monotonic in the cost of privacy. In fact, it is
17 often not monotonic. And that means that maybe there
18 is some optimal cost of privacy.
19 That led me to another question. What if we
20 could look at firms that need data in order to service
21 consumers, say, banks, lenders? And with those firms,
22 even in a competitive setting, would they collect an
23 appropriate amount of information or would they
24 collect too much? Even if they had no reason to
25 collect other than to service the consumers, not to

1 offer them other products but just to sell them one
2 product. And the result was that they collect too
3 much, and why do they do that? Well, because they
4 want to offer lower prices. And how do they offer
5 lower prices? By better fitting the consumer to the
6 product. So even in a market where data has no value
7 other than to screen consumers, too much ends up being
8 collected.
9 And that brought me to the next question.
10 What if firms could -- sorry. Wrong button. Wrong
11 button. It just keeps going. Further back. Okay. I
12 guess these slides are not there. It is okay. The
13 panel slides? That is all right.
14 The next model was one where those lenders
15 could actually sell the data downstream. They could
16 sell it to, say, insurance sellers. There we go. And
17 in those cases, firms actually collected even more
18 information. Okay? Now, is that good or bad? We
19 took the model to the data and the result was that
20 that could actually benefit consumers. Specifically,
21 we looked at five counties in the San Francisco
22 metropolitan areas. Three of those counties adopted
23 an opt-in approach, where you cannot sell consumer
24 data unless the consumer explicitly gave you the
25 consent do so. And the two other counties,

1 specifically the County of San Francisco and Marin,
2 had to opt-out approach where they could sell consumer
3 data unless the consumer actively opted out.
4 It turns out most consumers just do not
5 bother. They just go with the default. So if the
6 default is that you need to give consent, you never
7 give consent. And if the default is that you need to
8 actively opt out, you never opt out. Okay? So
9 effectively, these two regimes resulted in a regime of
10 privacy and a regime of no privacy. All right? One
11 where your data could be sold and one where it could
12 not.
13 Now, when your data could be sold, prices
14 were lower. And in the downstream, there were less
15 foreclosures. So in some sense, consumers were better
16 fitted with financial products. So here we see, sure,
17 we might like that our data cannot be sold without our
18 explicit up-front consent, but there are costs to
19 that. Costs might be we pay more. The other cost
20 might be that we are more poorly matched with
21 products.
22 So that led me to a bunch of other models
23 where I wanted to see what happens if we cut off
24 firms' access to consumer data. And those are widely
25 spread models. Those are models that I used in

1 antitrust cases, for example. And I looked at the
2 results for each of these in terms of consumer
3 surplus, firm profit, whether some consumers prefer
4 privacy or not, and overall welfare. Now welfare in
5 the sense you pay more, you pay less, welfare from the
6 perspective of prices.
7 So interestingly enough, in almost all of
8 these models, consumers were actually worse off in an
9 overall sense when their data could not be used to
10 target offers to them. Now, of course, there is no
11 intrinsic benefit to privacy modeled here. This is
12 all about prices. Now, firms actually could benefit
13 because the restriction not to sell data acted as some
14 sort of a solution to this prisoner's dilemma where we
15 are competing on fewer fronts now. It actually led to
16 higher profits.
17 The next question with this model was what
18 if we are looking at a merger case where, say, we have
19 three firms in the market and two of the three are
20 potentially merging? What would happen to consumer
21 surplus in this case if, on the one hand, firms could
22 access data and on the other they could not? And the
23 result was kind of not what we expected. Okay?
24 Merger policy turned out to be even more lenient when
25 firms could access data. It was easier to approve the

1 merger when firms had access to data.
2 And the reason, again, was that firms
3 competed on all these fronts when they had data. They
4 could segment the population where that led to more
5 competition and that resulted in lower prices which
6 increased consumer surplus. Okay?
7 So we tried to extend this. We looked at a
8 variety of market structures. You can think about
9 firms being spread in terms of consumer tastes and
10 some firms may have more customers buying from them.
11 Others not. And if we think about firms A and B
12 merging in this context, then the picture on the left
13 depicts the cases where consumer surplus actually does
14 not suffer much as a result of the merger.
15 Specifically, those areas that are shaded dark
16 basically represent market structures where it would
17 be easy to approve the merger because of the fact that
18 firms have access to data. Okay? So data does
19 influence or should influence merger policy.
20 So this brings me to the final topic that I
21 will discuss later today, as well. We just recently
22 started looking at the effect of the general data
23 protection regulation in the European Union on
24 investment and technology ventures. So if you look at
25 these two figures, the top one shows the average

1 amount in millions of dollars invested per deal in the
2 European Union and in the U.S. The U.S. is the orange
3 curve. The European Union is the blue curve.
4 And you can see that they more or less track
5 each other somewhat well up until GDPR takes effect in
6 May of this year, and things start to kind of diverge
7 a little bit. If you look at the second graph, it
8 looks at the total number of deals, venture deals.
9 Think about seed rounds, series A, series B rounds,
10 and so forth. All of those deals were technology
11 ventures and raised money. You can see that again
12 after GDPR, they started to diverge again.
13 So we could look at this difference and try
14 to quantify it a little bit and see what the impact is
15 on those firms and the result is quite significant,
16 that those firms begin to raise less money. And fewer
17 of those firms come to fruition because there are
18 fewer funding deals. So the regulation has a
19 noticeable impact. Now, of course, we do not know
20 whether this is a long-term impact or whether this is
21 just a short-term reaction. We only have several
22 months of post-GDPR data. But it would be interesting
23 to find out.
24 At least from the short-term perspective, we
25 can see that there is a significant impact. And this

1 impact can translate into an impact of the products we
2 see, maybe some products do not come to fruition.
3 Maybe those products are developed within established
4 firms entrenching their market power. Maybe some of
5 those products should not come to fruition. Maybe
6 they are bad products, products that abuse our data,
7 and this regulation is helping prevent that. We do
8 not know the answers to that. But what we can see is
9 that less investment has taken place. And we can
10 translate that reduction in investment into an effect
11 on jobs.
12 And we can see from our calculation that,
13 for firms that are relatively nascent, those are new
14 firms, they are about zero to three years old, the
15 amount of dollars they raise per employee is somewhere
16 between $120,000 and $1 million. Okay? And we can
17 translate that into a very rough preliminary range on
18 the potential number of job losses that they incur as
19 a result of GDPR, somewhere between 3,000 and 30,000
20 jobs. And as a fraction -- as a percentage of the
21 amount of employment those firms retain at least in
22 our sample, it is substantial. It is between 4 and 11
23 percent.
24 So just some overall observations that we
25 have also seen in the literature here, we have

1 theoretical papers that show that identical compliance
2 costs with data regulation tend to disproportionately
3 burden smaller firms. This is something that we saw
4 with the rollout of GDPR. We do not know if it is a
5 long-term effect, but at least in the short term.
6 Another result shows that compliance costs
7 can push innovation into happening inside established
8 firms. This is also somewhat confirmed by what we see
9 at least in the short term. And some final
10 observations here, it seems that any regulatory
11 approach should embrace nuance. It should be dynamic.
12 It should be market and context-specific. If we just
13 have a blanket approach, we are just likely to burden
14 smaller businesses and maybe entrench market power.
15 Now, using data regulation, data privacy as
16 kind of a means for data security is intuitive. It is
17 something that makes sense. But we should strike a
18 proper balance. We should not prevent altogether the
19 use of personally identifiable data just because it
20 makes it easier to have data security. Okay?
21 And then, finally, we should incorporate
22 data considerations into merger review because we see,
23 at least in our models, that they do have an effect.
24 Thanks very much.
25 (Applause.)

1 MR. SANDFORD: Thank you, Liad.
2 Our final presenter will be Florian
3 Zettelmeyer.
4 MR. ZETTELMEYER: Thank you. Well, thank
5 you very much for having me here. I appreciate the
6 invitation very much.
7 I am going to talk about a topic which is
8 quite different than what our prior speakers have
9 done. I am going to sort of take the perspective of
10 what it is that we, as observers, could learn about
11 what is going on. In other words, both as academics
12 but also inside firms. And as a result of that, the
13 basic thesis that I am going to propose to you today
14 is that firms are increasingly adopting machine
15 learning in order to do advertising promotions,
16 inventory optimization, whatever it is to basically
17 run their business.
18 In many cases, these things now are
19 colloquially interpreted as being AI, a term that you
20 might have heard, which is, in practice, not well-
21 distinguished from machine learning. And the point
22 that I am trying to make is that these
23 high-dimensionally targeting algorithms that exist out
24 there are creating very, very strong selection
25 effects, which make it very difficult to use

1 traditional measurement methods in order to kind of
2 disentangle what happened and what was going on.
3 And I want to give you an example of a study
4 that I have done and then I will talk to you a little
5 bit through where I think some of these problems are
6 coming from. So I ended up -- for today, the study I
7 want to refer to is the following question, which is
8 that -- so you may be aware of this that there the
9 most overused quote in marketing ever is a quote by a
10 guy called John Wanamaker that says, "I know that half
11 of my advertising is wasted. I just do not know which
12 one, which half."
13 And this was something that had a lot do
14 with the way that firms have traditionally been able
15 to track advertising measurement, and the way they did
16 it is that, you know, you basically had maybe a sense
17 of how many people you reached with an ad, so think of
18 TV advertising 40 years ago, and you had kind of a
19 sense of how many people bought. But you could not
20 link at the individual level who bought and who was
21 exposed to any kind of advertising.
22 So what happened over the last 15 years or
23 so is that this link is now possible. We know in the
24 case of Google, in the case of Facebook, in the case
25 of many of the advertising platforms, we can typically

1 track who ended up being intended to be targeted with
2 an ad, who actually got targeted, did they click and
3 then did they purchasing something as a result?
4 So the question that we have for us was
5 originally motivated by an industry concern not by a
6 regulatory concern is, does great data with
7 observational nonexperimental methods as are common to
8 user industry allow you to basically accurately
9 measure advertising effects? That was the basic idea.
10 Now, what we did is we ended up teaming up
11 with Facebook to answer, with a marketing science
12 group at Facebook. And they had just introduced, when
13 we started this project a few years ago, a product
14 called a Facebook "Lift Test" tool, which was a tool
15 to run randomized control trials within the Facebook
16 platform. This turns out to actually be a very
17 difficult thing to do.
18 You will hear tomorrow from another
19 gentleman, Garrett Johnson, who can tell you how hard
20 it was to implement this at Google as well. There
21 were a lot of technical details about how to make
22 experimentation work in these settings in which
23 algorithms are essentially -- they are sort of
24 machines to break probabilistic equivalents that you
25 need for testing.

1 And in this case, we looked at 15 large-
2 scale RCTs across a number of different industries.
3 We chose them. They were not supposed to be
4 representative of Facebook advertising. We chose them
5 because they were large enough sample sizes and we had
6 good outcomes we could measure, et cetera. We had
7 about between 2 and 150 million users per experiment,
8 over 1.4 billion ad impressions.
9 You have to understand that the Facebook
10 data is unusually clean because of the fact that
11 Facebook requires a single-user login which means that
12 you do not have any problems about misidentifying
13 people because their cookies do not match up. And we
14 ended up measuring real outcomes. Most of them were
15 real purchases; in some cases, registrations or
16 website views. But it was mostly actual purchases at
17 online retailers.
18 Now, you also have to understand that we
19 were able to measure what people did even if they did
20 not click on anything, because of the fact that we
21 could later trace who had been exposed to an ad to
22 that consumer's identity back at the advertiser. Of
23 course, we had no personally identifiable information
24 about any of these people.
25 So let me give you an example of this study.

1 So here is a study that was 25.5 million users. Think
2 of this as like an ecommerce website where you can
3 purchase something online. Thirty percent were in the
4 control group; 70 percent were in the test group. The
5 outcome of the measuring was this purchase at a
6 digital retailer. You have what is called a
7 conversion pixel, which the advertiser placed after
8 the checkout page. So this study ran for 17 days,
9 which is a pretty normal duration.
10 So what we then do is we measure the lift
11 from the randomized control trial sort of to establish
12 a ground truth. And the basic issue here is that in
13 advertising, you cannot guarantee that anybody is
14 exposed to an ad, so these kinds of experiments always
15 intend to treat designs. In other words, you can say,
16 I would like to expose you to an ad, but whether you
17 actually see the ad depends on many things. Like are
18 you trekking in Nepal or are you logging into Facebook
19 today or whatever it is or maybe -- you know, maybe
20 somebody else kind of bid for your ad impression. As
21 a result, you did not get to see the ad.
22 And so in -- let’s say as an example in our
23 situation, we had about 25 percent exposed user, 75
24 percent unexposed users and we had a control group
25 that we could guarantee was unexposed. Okay?

1 So in this particular case, what we did is
2 using sort of traditional average treatment effect on
3 the treated, we observed a conversion outcome of 0.104
4 percent in the exposed group and then calculated a
5 counterfactual conversion outcome in the control group
6 of 0.059 percent. So these are users who would have
7 been exposed if they had been in the test group.
8 And what this tells you is that -- and this
9 is the traditional way that a company would express
10 this -- there was a lift of 73 percent. So as a
11 result, sales increased by 73 percent due to the ad.
12 Okay. So think of this as kind of the gold standard
13 truth running through a randomized control trial.
14 So, in practice, what now happens is that
15 many advertisers do not use control groups. In fact,
16 this is the norm. It is relatively rare to run
17 randomized control trials. So, in our situation, what
18 we basically had is a situation where, since our
19 testing control groups are randomly assigned, we could
20 replicate what you would -- the situation you would
21 find yourself in as an advertiser if we just threw
22 away the control group and just operated with a test
23 group as being our group where we could see that some
24 people were exposed versus unexposed.
25 In this particular case, it turns out that

1 if you then compared the probability that somebody
2 purchased in the exposed versus the unexposed group,
3 the actual measurement of how well somebody did, in
4 other words, we take people who saw an ad, we took
5 people who did not see an ad, all of which were in the
6 target group, in the test group, the measurement of
7 how well the ad did went up to 316 percent. In other
8 words, a massive overestimate of how well the ad is
9 actually working.
10 Okay. And so it turns out, of course, that
11 the fact that you get biased measurement because
12 exposure is endogenous in this industry is well known,
13 and as a result of that, a lot of ad measurement
14 companies like, for example, comScore that I have
15 listed here on this example slide from one of their
16 decks, basically says, what we are going to do is we
17 are going to take an ad-exposed group and then we are
18 going to have test and control groups that are matched
19 on demographics and behavioral variables, which gives
20 us a balanced unexposed group, which sometimes is
21 referred to in this industry as a forensic control
22 group. So one that you create exposed using matching
23 methods and things like that.
24 Okay. So what we did is we said, okay, we
25 have pretty good data, because at Facebook, there is

1 great data about what consumers do. Let us see if we
2 could actually replicate a good balance unexposed
3 group that would allow us to measure what is going on.
4 So we tried. So the basic idea is that we are taking
5 people in the exposed group and then we are taking a
6 subset of the people in the unexposed group that by
7 anything we observe about them should be somehow
8 equivalent to the people in the exposed group.
9 Good. So in order to do this, we use the
10 best of what exists in industry and academia, at least
11 at the scale that we use, there are more sophisticated
12 methods, but they do not work with 150 million users.
13 So we used exact matching, propensity score matching,
14 stratification, regression, inverse probability-
15 weighed regression adjustment, stratification and
16 regression, and we had really wonderful data because
17 we have data on Facebook characteristics and,
18 moreover, we even have data on -- Facebook ends up
19 having an internal algorithm where you, as an
20 advertiser, give Facebook a set of email addresses and
21 then say, find me other users at Facebook that are
22 like the users that are represented in these email
23 addresses but are not these users.
24 And what we used is we literally used their
25 algorithm to do this, which is a massive machine

1 learning based algorithm in order to find a balanced
2 unexposed group for the exposed group. Okay. So in
3 other words, we threw at it what is really unusually
4 good data in order to do this.
5 So let me show you the result. So what you
6 see up here is the following. You see that the
7 benchmark lift is 316 percent. That is what we found
8 from the exposed-unexposed measurement. The benchmark
9 in the RCT is 73 percent, which we take to be the
10 truth. And what you now see here is essentially a
11 sequence of methods that end up -- you notice there is
12 sort of stratification and then propensity score
13 matching and regression, et cetera, that end up
14 becoming better and better as you add more data. So
15 every method is essentially there were three or four
16 variable sets.
17 And you notice in this case, the world looks
18 hopeful because you can approximate pretty well with
19 the normal observational methods. So you, as an
20 advertiser, could do this or we, as a researcher,
21 could do this. More or less, what happened in the
22 industry. Well, the problem is -- and then so you do
23 this on another method and it looks wonderful. Like,
24 there seems to be a consistent pattern across methods
25 and you start feeling very hopeful about the ability

1 of recovering with the data that we normally observe
2 what our cities do, until you hit one of the other 15
3 studies, and suddenly it looks like this.
4 The truth is 2.4 percent. And the closest
5 estimate we have is a 1306 percent lift. So this is a
6 study, by the way, where only 6 percent of consumers
7 actually got exposed to the ad. And what that means
8 is that there was a huge amount of ability for the
9 model of essentially targeting those consumers and
10 making them very different from the unexposed group.
11 You also sometimes find when you get used to the idea
12 that maybe there is always an overestimate, that
13 sometimes these methods actually totally underestimate
14 what is going on.
15 Good. And so for me -- sorry. I should
16 have warned you about this. Red means massive
17 overestimate. White means more or less okay. Blue
18 means underestimate. And you see that it is all over
19 the map depending on the studies. And so it is very
20 difficult for us ahead of time telling you what is
21 going to happen without knowing more about these
22 particular studies.
23 Okay. So the basic idea is this, which is
24 that we are in a situation and it is because of the
25 fact that firms are using machine-learning models,

1 where the targeting of consumers is becoming so
2 basically deterministic that a lot of the
3 observational methods that we use, which rely on the
4 idea that there remains random variation after you
5 condition out what we observe of people, start
6 breaking down. And this is quite important because it
7 means that this lack of transparency that Alessandro
8 was talking about earlier is all over the place.
9 So you have an industry where, for example,
10 many people who spend a ton of money on marketing at
11 the moment simply do not know how well these kinds of
12 interventions are working, because unless you plan
13 ahead big-time and spend lots of money on doing
14 randomized control trials, you literally have no sense
15 of being able to tell whether your expenditures are
16 actually working or not. And this is important both
17 -- so it is this really interesting thing where
18 despite amazing data -- and these algorithms make it
19 very difficult to actually get accurate feedback on
20 what is going on in industry.
21 And this is not well understood in industry
22 and it creates sort of a level of grayness that I
23 think a lot of people do not expect in this particular
24 industry. Thank you very much.
25 (Applause.)

1 MR. SANDFORD: Okay. So once again, if you
2 are in the room here at American University and would
3 like to ask a question, we have people walking up and
4 down the aisles with note cards. So please flag one
5 of them down.
6 Okay. So my first question is -- I am
7 looking back at Joe Stiglitz's remarks from a prior
8 hearing and he opined that big data provides new tools
9 for price discrimination and those with ability to
10 discriminate better grow. And so the firms that get
11 big and become successful are those with lots of data
12 that can do price discrimination and not necessarily
13 those with the best product.
14 And Liad's presentation talked about the
15 effect of privacy in a price-discrimination context.
16 I read a survey paper by Alessandro Liad and Curtis
17 Taylor and many of the papers there talked about price
18 discrimination, again, as sort of the vector for how
19 privacy affects consumer outcomes. And the question I
20 have here, you know, 20 years ago, I would have said
21 it was obvious that we were headed for an era with
22 individualized pricing. I would go on Amazon and I
23 would get a price that only applied to me. Indeed, I
24 wrote a paper for my intermediate microeconomics class
25 saying as much. The paper received a B for good

1 reason. It was completely wrong. That did not
2 happen. It has not come to pass.
3 I have a quote from computer scientist,
4 Arvind Narayanan. He wrote "The mystery about online
5 price discrimination is why so little of it seems to
6 be happening." And so from my perspective, the price
7 discrimination I do see online is the same thing that
8 retailers were doing 100 years ago. It is coupons and
9 sales and starting the price high and then lowering it
10 over time.
11 So my question is, why don't we see more
12 price discrimination? And if you agree with my
13 premise that we do not see a lot of price
14 discrimination, should that cause us to update our
15 priors of how we think about privacy given all of this
16 work on the effect of privacy on welfare through price
17 discrimination? So I will just throw that out to the
18 panel, anyone who wants to answer it.
19 MR. WAGMAN: I would just like to say --
20 MR. SANDFORD: You have to turn your mic on,
21 by the way.
22 MR. WAGMAN: I would just like to cite a
23 couple of examples that we are starting to see
24 individualized pricing, at least in the context of the
25 ridesharing apps where the price you see is very, very

1 likely tailored to your record, your history of using
2 the app, for example, on Uber or Lyft. And those
3 efforts are, in my estimation, only intensifying over
4 time.
5 The other piece of what you mention of
6 offering coupons and discounts, I think that can also
7 be a lot more targeted than it used to be. And so we
8 may not see price discrimination upward from, say, a
9 certain base price or a perceived base price, but we
10 will certainly see it downward with only certain
11 selected individuals being targeted with offers.
12 MR. SANDFORD: Ginger?
13 MS. JIN: I just want to add that, from the
14 economic point of view, the word "discrimination" is
15 probably not as loaded as it sounds in plain English.
16 According to this theory, price discrimination is not
17 necessarily welfare-reducing whether that is defined
18 for consumer welfare or total welfare, because when
19 you are comparing with uniform pricing, when you have
20 price discrimination, some people may get a discount
21 from that and some people may have a price higher than
22 the uniform pricing. So the welfare consequence of
23 that is going to be a mixture depending on how many
24 people are getting a discount and how many people are
25 getting a lift.

1 And in terms of underuse of data in price
2 discrimination, I think there is still some probably
3 preference -- consumer preference about sort of to
4 what extent the firms are using price discrimination.
5 I think that is probably a separate dimension as
6 compared to sort of their willingness to pay for a
7 particular product. And if a firm has a sense that
8 consumers dislike this kind of personalized price
9 discrimination, even if they make a short-term
10 discount on this particular product, I think a firm
11 will take that into account.
12 I mean, this is just hypothesis. I wonder
13 to what extent that kind of general resistance to
14 personalized price discrimination actually get into
15 firms' sort of choice of how much price discrimination
16 they would use.
17 MR. SANDFORD: Florian and Alessandro?
18 MR. ACQUISTI: Thank you. Echoing something
19 Liad was saying and connecting it to what Ginger was
20 saying, I believe that part of the puzzle is that what
21 may be happening is something I call product
22 discrimination. And as Ginger pointed out, I am not
23 using the term "discrimination" with a negative
24 connotation but rather in the economic connotation.
25 By product discrimination, I am referring to the

1 ability of the industry, the advertising industry, to
2 send an ad for a certain product rather than another.
3 In doing so, they may match a consumer to what is
4 maybe a higher-quality or a lower-quality product,
5 higher price, lower price.
6 So we may not see the very same product
7 being sold at different prices to different consumers.
8 So we may not see first-degree price discrimination,
9 which is most of what the empirical efforts have been
10 trying to do. But we may see basically forms of self-
11 selection, second-degree price discrimination.
12 By the way, one very small pushback, I would
13 contest the notion that much of the negative welfare
14 consequences of privacy for consumers are related to
15 price discrimination. That is one part of the story,
16 but there are others.
17 MR. ZETTELMEYER: You know, I think another
18 aspect of price discrimination is that we have -- the
19 question of exactly what is price discrimination, what
20 is intertemporal pricing is actually not very well
21 defined. A nice example of this -- and I will tie
22 this back to online markets in a minute. A nice
23 example of this is I have done a study on pricing at
24 car dealerships, and it turns out you can actually
25 explain a lot of the -- what looks like price

1 discrimination, namely different consumers are paying
2 different amounts of money for the car, simply by the
3 levels of inventory that happen to exist when the
4 consumer is walking into the dealership.
5 So it looks like the dealer is
6 discriminating against individual consumers, but it is
7 really reflecting the scarcity rents of the inventory
8 that happens to be lying around. So if you have two
9 red Honda Accords on the lot, you are going to price
10 it differently than if you happen to have 53 on the
11 lot. And depending on when you walk in as a consumer,
12 you are going to see different prices.
13 To us it looks very similar, as if it is
14 price discrimination, but it actually has a very
15 different economic reason for it. So I think
16 similarly in the online context, you do observe a lot
17 more intertemporal price variation and we can think of
18 that as also being at least, you know, fulfilling a
19 similar goal as individual level price discrimination,
20 first-degree price discrimination literally at the
21 same time for the same kinds of consumer.
22 So I guess there is maybe more price
23 discrimination than meets the eye, which I think was
24 Liad's point as well.
25 MR. BEN-SHAHAR: I guess I will add my

1 perspective. I think that the puzzle is compounded by
2 the fact that we do not see personalizing of other
3 aspects of the product not just the price. Why should
4 everybody get the same right to return the products,
5 the same warranty, the same privacy terms? If we know
6 enough about people, how much they can pay, we
7 probably know a little bit, also, or a lot about what
8 their preferences are.
9 We do see that, you know, people are
10 sometimes thrown out of Amazon Prime if they are
11 return-aholics or things like that. So it is either
12 zero or one, but we do not use a dimmer and that is
13 kind of puzzling to the same extent that the -- now, I
14 guess one of the problems that jumps to mind, and I
15 have not studied this closely with the data, but is
16 the problem of arbitrage. As long as you are selling
17 products and not services, people can resell them.
18 I think that once things are done through
19 platforms, apps, and are sold as utilities and
20 services, we might be able to -- we might see the
21 burst of a personalization of various aspects of
22 products.
23 MR. SANDFORD: Okay, thank you.
24 Next question, so Omri mentioned the privacy
25 paradox which is, as I understand it, is consumers say

1 overwhelmingly that they prefer greater privacy, yet
2 they do not act in a way consistent with that. So,
3 for example, I think I pulled from Alessandro and
4 Liad's paper, 86 percent of your adults say they do
5 not want targeted advertisements; 93 percent of all
6 adults believe in "being in control of who can get
7 information about them is important." And, yet, it is
8 not clear that consumers behave in a way consistent
9 with the preferences expressed in surveys that ask
10 you, yes or no, do you prefer greater privacy.
11 So, I mean, my reaction to this is -- well,
12 one, is this actually a paradox? I mean, is this just
13 we are suggesting something that sounds vaguely
14 positive to people and saying, are you in favor of it
15 or not and they say, yeah, sure, I am in favor of
16 animal rights but I like to eat steak. I mean,
17 something like that.
18 And two, I mean, kind of -- is there in a --
19 you know, we look at firms in the market. They have
20 different privacy policies. Is there a sense in which
21 consumers have different preferences over these
22 different privacy policies and might go to one firm or
23 another based on their privacy policies? So do
24 consumers have a downward sloping demand for privacy
25 that is -- you know, has meaningful slope across the

1 range of privacy policies we see in the marketplace?
2 Whoever wants to go first.
3 MR. ACQUISTI: I may start. I feel that
4 there is quite substantial evidence that there is a
5 demand for privacy by consumers and this demand
6 follows, to some extent, canonical, traditional,
7 expectable economic lows. People will exercise their
8 demand for privacy when the price of doing so is
9 small. People close their bathroom door when they are
10 going to the bathroom. People do not post their
11 credit card online because it would be insecure and it
12 would be also probably costly, just the act of doing
13 so.
14 As you get into more esoteric and costly
15 behavior, consumers engage into that when there is an
16 actual benefit for doing so. So wealthy individuals
17 go to quite extreme measures perhaps sometimes to hide
18 their wealth and use bank accounts which may not be
19 monitored by enforcement agencies, for instance. And
20 they try to have anonymity and they may pay for that
21 because it is very valuable to them. So there is
22 actually a demand for privacy which follows canonical
23 economics lows, but there are also these issues of not
24 always being able to predict what the cost of privacy
25 will be especially online for -- due to the fact that

1 privacy tradeoffs are intertemporal in nature.
2 So you may reveal information now which may
3 not affect you for a long time, but eventually will
4 affect you. And this, to me, one of the possible
5 explanations, not the only one, for the privacy
6 paradox.
7 What is very interesting to me and Omri made
8 me think about that through his remarks is that there
9 is another form of paradox which is much less explored
10 but as compelling. The paradox of people who claim
11 that privacy is not important to them, but, in fact,
12 act as it is. And that is really many of us. Even
13 though the people who claim that privacy is not
14 important engage in behaviors every day, both online
15 and offline, which are privacy-seeking behaviors,
16 lowering the tone of the conversation in the
17 restaurant when they are having dinner with their
18 partner when the waiter arrives. That is a form of
19 privacy-seeking behavior in public where you are
20 trying to make your conversation private.
21 The example I was making earlier of closing
22 the bathroom door when you go to the bathroom; the
23 other example I was making earlier of not sharing your
24 credit card information online. Now, if you ask
25 people about these behaviors, some would probably, in

1 a manner, suggest that it is not about privacy. For
2 instance, it is -- not sharing the credit information
3 online is about security. Closing the bathroom door
4 is about social norms or politeness, not privacy. To
5 me, this suggests that people have very personal
6 definitions about what privacy is, and it is not an
7 intent to disregard other people's definition of
8 privacy in favor of their own. But, in fact, at the
9 end of the day, they are all about the same thing,
10 which is the individual's ability to modulate the
11 degree of public and the private in their lives.
12 MR. SANDFORD: Ginger?
13 MS. JIN: Yes. I just want to echo
14 Alessandro that there is a definition problem here.
15 If we think privacy protection or data policy is one
16 product attribute for the product and service I am
17 buying, it is unclear exactly what is that product
18 attribute I am buying. Okay? So you can think of,
19 say, 100 percent protection on one end and zero
20 protection on the other end. I am actually not sure
21 exactly where I am buying in that spectrum because the
22 firm may protect my data very well or run with it.
23 Right? So we do not know exactly. And that fuzziness
24 probably could be one of the explanations for this.
25 Another related issue I want to echo was

1 Omri's point about data pollution. I think from
2 consumers' point of view, if you view data policy as
3 one product attribute but you just do not have time to
4 track exactly where that product attribute is for
5 every firm, every product you are having, you have
6 this overall impression. Okay? And then when you
7 heard about Equifax or Cambridge Analytica or
8 something, you sort of formed this kind of prior or
9 posterior about exactly where this product attribute
10 is. And that is evolving.
11 And it could be this firm actually doing a
12 very protective thing about my PII data, but because I
13 heard so many other things that I got sort of afraid.
14 I am afraid you are going to run with my data for some
15 abusive use. So in that sense, you probably get to
16 the second paradox that Alessandro was just talking
17 about, which is how can I convince you that I am
18 actually selling you a product with a very good data
19 policy? It will be very hard to convince given that
20 your prior is sort of polluted by many other firms.
21 MR. SANDFORD: Florian?
22 MR. ZETTELMEYER: Yeah, I think to tie some
23 of these things together is simply the link between
24 data and what is done with it is so opaque today, and
25 I think that is what is leading to a lot of the

1 problems. So exactly the same data could be used for
2 ways that absolutely delight you and then for ways
3 that you would find absolutely horrendous. And so I
4 sometimes wonder whether we spend too much time
5 thinking about how to protect the data as opposed to
6 protecting the use of the data. And I think, you
7 know, in some sense, it is the entire promise of this
8 big data enterprise. And if you think about the
9 current advances in machine learning, it is that data
10 can be used in ways that should blow all our minds in
11 order to form predictions that we never thought could
12 reasonably form with data like that.
13 And as a result, somehow being able to
14 expect that people can have reasonable agency with
15 regards to the protection -- what data they make
16 available in the complete lack of a link between what
17 happens with their data and -- between them giving
18 their data out and what happens with their data is
19 incredibly difficult to accomplish. It is like asking
20 somebody to regulate the electricity usage at home if
21 they have absolutely no idea what the usage of any
22 device is and they cannot measure the outcome of it.
23 How do we expect people to be somehow reactive to how
24 much energy they are using? It is a very similar
25 situation in this realm as well.

1 MR. SANDFORD: Liad?
2 MR. WAGMAN: I think there is also a sense
3 of no matter what I do, it is going to be collected.
4 Just to give an anecdotal recent example, GDPR rolled
5 out and a large firm with millions of users put the
6 consent popup on their page. So when users would surf
7 to the page, they would see the consent. And they
8 would have two options. They could say, yes, I am
9 willing to share everything, or, no, I want to choose
10 what I share. 96 percent of users clicked on yes, I
11 will share everything. And 4 percent clicked on, no,
12 I will choose what I share.
13 And then they clicked on that and they very
14 carefully chose -- they had the option to choose to
15 share nothing. But they very carefully chose to share
16 some and not others. And interestingly enough, based
17 on their choices, they could be easily identified and
18 targeted with ads, because their choices were highly
19 correlated with other information about them. And so
20 there is this sense of inevitability, no matter what I
21 do, it will be collected and I will be identified at
22 least in some sense.
23 MR. ZETTELMEYER: Or worse, actually,
24 machine-learning algorithms are going to figure out
25 what my preferences are even if I do not state them.

1 MR. WAGMAN: Right.
2 MR. BEN-SHAHAR: I would like to touch on
3 two things that the panelists said. I would like to
4 challenge Alessandro's response. He said, you know,
5 people close the bathroom doors. You see there is
6 privacy. You know, but they do not mind the
7 electronic eye that flushes the toilet. Right? Even
8 if there was...
9 (Laughter.)
10 MR. BEN-SHAHAR: I mean, that is, I think,
11 the difference between the privacy -- the secrets that
12 we have in the presence of other people and the data
13 privacy, vis-a-vis, the algorithms that are collected.
14 You know, even if the electronic eye was connected to
15 some algorithm and sold me some constipation
16 medication, you know, I think people initially might
17 be alarmed. But, ultimately, I think it would not be
18 out of a -- it would not change their behavior to use
19 these bathrooms pretty comfortably.
20 So I think that you probably have a lot of
21 evidence that people care about data privacy. I would
22 not use the example of closing bathroom doors to make
23 that -- that seems a little bit like kind of a
24 strawman.
25 I really like the point that Liad made that,

1 you know, look, four people -- only 4 percent of the
2 people exercised what a lot of privacy advocates and
3 privacy regulators want them to, which is user
4 control. I actually think that 4 percent way, way,
5 way overestimates the prevalence of this phenomenon
6 once the novelty will die out and we will realize that
7 you have to do this not to that one website in that
8 experiment or whatever, but to do it to dozens of
9 places daily and that you really do not know what are
10 the right choices because you do not know what the
11 tradeoffs are. You do not know. It is so
12 complicated.
13 User control in every aspect -- I have
14 studied that not in the privacy context but in
15 consumer credit, probably a much more fateful decision
16 people make -- user control is kind of a panacea.
17 People cannot make good decisions no matter how well-
18 intentioned regulators are to give them all the aids,
19 decision aids and choice architecture if they do not
20 understand the tradeoffs and they do not have the
21 sophistication to deal with problems that, at the
22 core, are not simple.
23 MR. SANDFORD: Alessandro, you wanted to
24 make a brief point?
25 MR. ACQUISTI: Very brief comment. I

1 actually do not disagree with you, but the contrast
2 between online and offline was intentional. It was to
3 point out that there are situations where individuals
4 take action to protect their privacy, especially when
5 it comes to physical privacy, and there are situations
6 where they may not, especially when it comes to online
7 privacy.
8 To me, from this to conclude that that
9 implies that people do not care about online privacy,
10 that is, to me, the conclusion that is erroneous,
11 because there are many, many factors which
12 differentiate the offline scenario, the bathroom door,
13 and the online scenario, including intertemporal
14 tradeoffs. You are seen immediately by someone else
15 in the bathroom. If you post something, you may not
16 be seen by someone who with an interest to use your
17 data one year later, five years later. They show
18 information asymmetry.
19 The issue that Liad was referring to of
20 efficacy, if I close the door, I have control. If I
21 post something on Facebook, even if I use correctly
22 the privacy settings and visibility settings, I still
23 do not have really much control on what happens to
24 that photo after I uploaded it. So it is intentional
25 for me to contrast the online and offline. As a

1 matter of fact, we do have a paper that is about to be
2 submitted about this in particular.
3 (Laughter.)
4 MR. SANDFORD: Okay, thank you. So it
5 sounds like obfuscation. It is not clear to me what
6 the privacy policy is and frustration with that is a
7 driver of why consumers do not seem to care about
8 privacy.
9 I wanted to ask Omri a question before he
10 has to leave. Omri, you wrote a book with Carl
11 Schneider espousing your view that privacy policies
12 are essentially worthless. No one reads them. You
13 said that, in 2008, it would take 76 workdays to read
14 all of the terms of use and privacy policies that one
15 would come across in the course of normal use of the
16 internet, and that was ten years ago. It could be
17 more than 365 workdays now for all we know.
18 Omri had a picture in the book where Omri is
19 like two inches tall in the photo and the iTunes terms
20 of service are like a foot tall in the photo. I mean,
21 they come down from the second floor and dwarf him.
22 So his point is it is effectively impossible to read
23 everything that you are agreeing to when you use
24 various websites, and so I do not want to put -- my
25 characterization on Omri's is that these are

1 essentially useless. They provide no bite. They are
2 not helpful to consumers in deciding which websites I
3 should patronize and which I should not.
4 So I guess my question, Omri -- you can
5 respond to that however you want -- but my question is
6 how many people need to read these for them to be
7 effective? So for example, if a government plaintiff
8 reads a privacy policy and says, hey, you are not
9 behaving in that way, is that meaningful to what kind
10 of privacy policies get promulgated in the
11 marketplace? If a journalist reads one of these
12 policies and says, hey, there is something kind of
13 funny in this policy, would that scare users away and
14 be a check on what goes into the privacy policy? So
15 what do you make of that view?
16 MR. BEL-SHAHAR: Thank you for raising this.
17 I think the good people at Carnegie Mellon read the
18 privacy policies and grade them for us. I do not
19 think many people go to PrivacyGrades.org. I know
20 occasionally a newspaper, The New York Times, calls to
21 ask me questions about the terrible things that
22 Facebook does, and I say, look, your app gets a lower
23 grade then Facebook. But, of course, The New York
24 Times is not the problem, maybe Facebook is. And so
25 what are these grades really telling us?

1 I guess my view about giving people
2 information so that they will make wiser, more prudent
3 choices, is failing everywhere. It is not a privacy
4 problem; it is a disclosure problem. It is a problem
5 with the regulatory technique. It fails miserably and
6 for a long time in consumer credit where it all was
7 invented, truth-in-lending and things like that. It
8 fails all over contract law, because anytime you click
9 "I agree," people put you through these meaningless
10 rituals of clicking these things, closing boxes
11 because contract law requires consent for all sorts of
12 things that otherwise would be a violation of law,
13 including the privacy terms.
14 But also all the disclaimers and all the --
15 yada, yada -- all that stuff, all the consent forms in
16 hospitals that people get, 17 pages of consent forms
17 to participate in human subject research, the evidence
18 is -- the mountains of evidence -- undisputed that
19 nobody reads it. That the people to whom it is given
20 cannot understand it if they were able to read it and
21 the issues, as I mentioned, before are too complex.
22 So I guess in the privacy context, what many -- in
23 many places, the solution that is proposed and in
24 other contexts, too, is to simplify.
25 Simplification is, I call it in my book that

1 you mentioned, is the deus ex machina. It falls from
2 the ceiling and it kind of solves the plot and
3 everything is good afterwards. But it does not.
4 Simplification, in every area that I mentioned, has
5 been tried for decades and failed, again, for the
6 reason -- and now I am saying it for the third time --
7 that you cannot really simplify the complex. When
8 things are complicated, you cannot just give people
9 red light/green light.
10 And so I do not know -- I cannot
11 conceptualize in my mind, in response to your
12 question, who will actually read and give consumers
13 the information that will be operational? Ultimately,
14 consumers, if they want to make more prudent choices,
15 should rely on the experience of people like them. So
16 ratings sometimes help them and, in many contexts,
17 they do. They could also be misleading. And it is
18 very important to protect, as a regulator, the
19 integrity of these aids that do not give people
20 information, but give them a good prediction of how
21 content they will be if they actually jump into the
22 experience of this product or service.
23 MR. SANDFORD: Ginger?
24 MR. JIN: Just to add on Omri's point,
25 suppose we have a sophisticated machine that

1 government or journalists can use to really squeeze
2 out all the information from those pages and tell a
3 very simplified, but fully informative, story to
4 consumers, I think it does not solve the following
5 problem, which is how can I be sure what you say is
6 exactly what you do and given that what you do is
7 evolving over time with new technology and so forth.
8 So I think that the second part of this
9 problem is really crucial. Otherwise, you can say
10 anything. Right? So how can we sort of check what
11 you said and then make sure that is consistent with
12 the policy given the amount of data policy and the
13 kind of firms that could use data? I think it will be
14 unfeasible for everyone to be checked in a precise and
15 timely way, and I think that is probably one of the
16 inherent problems in this approach.
17 MR. SANDFORD: But there is still a
18 deterrent effect if there is a data breach that is
19 very high profile and that might get punishment from
20 the Government or something like that, or -- so
21 enforcement is sporadic, but perhaps severe when it
22 does come. That can still be a check on behavior,
23 could it not, on what goes in privacy policies?
24 MS. JIN: Well, when we talk about data
25 breach, it is a symptom, right? I mean, the agency,

1 like a doctor trying to come up with a diagnosis.
2 Unfortunately, the link between the symptom and the
3 diagnosis is not that straightforward. If a firm got
4 data-breached, it could be the firm's fault not having
5 enough security, so that it sort of left room for the
6 hackers to come in. Or it could be somehow the
7 hackers have the most cutting-edge technology that
8 will be able to penetrate even the most secure walls.
9 I mean, you have to tell those two in order
10 to say exactly is that a problem, the hacker's
11 problem, or is that a problem of the firm's problem?
12 MR. SANDFORD: Okay, thank you. Let's talk
13 now about supply for privacy. How do firms decide
14 what goes into their privacy policy and, in
15 particular, is there a sense in which firms are
16 responding to consumer preferences over privacy? We
17 have talked about how strong those preferences are,
18 whether they are reflected in consumer decisions or
19 not.
20 Do we see any evidence that firms are just
21 going for the maximalist privacy policy? I am going
22 to just write down everything I want and get you to
23 agree to it, or is there some sense in which firms are
24 responding to consumer preferences, maybe perhaps
25 worried that if I have a maximalist privacy policy,

1 users might shy away from my website and go somewhere
2 else for example?
3 So is there a sense in which there is a
4 supply curve for privacy that reflects firms either
5 giving a greater level of service in return for more
6 privacy or responding to consumer preferences for
7 privacy?
8 MR. WAGMAN: I think that is a tough one,
9 because those privacy terms are ever-changing. Right?
10 And if a firm realizes there is way to commercialize,
11 monetize, do something else with data, they will
12 change their terms so they can collect that data as
13 well. They might give some disclosure that, again,
14 nobody will read that they changed their terms. And
15 so I think it adds to that sense of, no matter what I
16 do, I cannot really prevent it being collected. And
17 even if right now the terms are friendly to me and
18 even if the firm actually follows through on those
19 terms, that can change at any time.
20 MR. ZETTELMEYER: Also, I am not sure to
21 which degree a lot of consumers understand the
22 difference between we have lots of your data and we
23 will keep it safe and we do not collect it in the
24 first place. Right? And so there are very few firms
25 that are using that from a branding point of view at

1 the moment. Apple is a very high-profile one. I
2 mean, as far as I can tell, I am not sure there is any
3 evidence that consumers necessarily care about that.
4 MR. WAGMAN: I would also add that even
5 those characterizations are sometimes misleading. So
6 even if a firm like Apple says, oh, we do not collect
7 it, they might have partnerships with other firms who
8 do collect it and they benefit from it indirectly.
9 MS. JIN: I think one anecdote probably does
10 suggest that people care at least about the perception
11 of privacy. I think that example was some years ago
12 Samsung had a TV and the TV has kind of a camera that
13 you can -- or voice recognition that you can sort of
14 give a voice command to the TV. And then there was a
15 kind of public outcry against the possibility that
16 maybe the microphone is always listening, even to the
17 private talks in your living room. And I think in
18 response to that public outcry, Samsung did change
19 their privacy policy. And again, does that exactly
20 reflected what they do in the future is still an open
21 question.
22 MR. ZETTELMEYER: But, of course, now we
23 have moved on to conversational interfaces like Alexa
24 that listen to everything you do and consumers seem to
25 be fine with it.

1 MR. ACQUISTI: Your point about a supply
2 curve for privacy is extremely interesting. It makes
3 me think about the -- another question that I find
4 under-explored in the research in this field, which is
5 the relationship between data collection usage of data
6 and the provision of free services and free content,
7 specifically to what extent increasing data collection
8 is necessary for the provision of more and better
9 services.
10 I know I am maybe about to say something
11 that sounds bold, but once again, I believe that I
12 have some empirical evidence to support the claim.
13 And the claim is that the relationship between data
14 collection and provision of services is more
15 correlational than causal or at least we do not have
16 very strong evidence of it being causal as opposed to
17 correlational.
18 What I mean is that the provision of free
19 services existed on the internet way back in the days
20 before the more granular techniques of collecting
21 information about users and tracking them across
22 different sites started, which is about 2004 or 2005
23 with Facebook, et cetera. Even nowadays, there are
24 firms which can do well without data collection.
25 DuckDuckGo is an example.

1 To me, this brings another question. Once
2 again, I really do not know what the answer is, but
3 the bold claim I am making is that I do not feel many
4 people actually know what the answer is, to what
5 extent the relationship between data collection and
6 provision of free services is correlational, to what
7 extent it is causal.
8 It goes back to the value allocation
9 question. To what extent when merchants may be paying
10 500 percent for targeted ads and publishers get 4
11 percent more for targeted ads. To what extent
12 something gets lost in the middle remains in the realm
13 of the data oligopolies. And this could potentially
14 provide an answer then to the question of causal
15 versus a correlational relationship between provision
16 of free services and data collection.
17 MR. SANDFORD: Florian?
18 MR. ZETTELMEYER: Can I ask you a question I
19 thought of, which is related to this issue? I think I
20 agree with you. I wonder, however, whether the one
21 exception to that is the current rise of AI and
22 machine learning in the sense that, if we think, you
23 know, roughly speaking as those being kind of
24 prediction machines that have large effects on the
25 quality of provision of services --

1 MR. ACQUISTI: And would not be able --
2 MR. ZETTELMEYER: -- and those cannot work
3 without data.
4 MR. ACQUISTI: -- exist without data.
5 MR. ZETTELMEYER: Exactly. So I think that
6 may be the one exception to that. And I am not quite
7 sure how to think through it, but I wonder what would
8 you think.
9 MR. ACQUISTI: I think you make a good
10 point. And it goes back then to an item I mentioned
11 at the very start of my talk, to what extent for that
12 kind of analysis we always need identified data versus
13 anonymized data, but to a degree of granularity, which
14 is sufficient for the kind of analysis. It goes back
15 to privacy not being monotonic, not being absence or
16 presence of data, but being a modulation of what type
17 of data you use and analyze.
18 MR. SANDFORD: Omri? Okay, thank you, Omri.
19 MR. BEL-SHAHAR: Sorry.
20 MR. SANDFORD: Okay. So the next question I
21 have is, is there a sense in which firms compete in
22 privacy policies or the answer may be no based on the
23 answers -- what we were just discussing. But, I mean,
24 is there a sense in which, you know, say Facebook has
25 a bunch of locked-in users that they can have a more

1 maximalist privacy policy than, like, Walmart that has
2 to go out and compete for every retailer dollar with
3 other online sites? And so is there a sense in which
4 competition matters for privacy? And is there a sense
5 in which, say, removing a competitor, like with a
6 merger, could matter for privacy outcomes?
7 And your answer can be no, in which case we
8 do not need a long -- it need not be long.
9 MR. WAGMAN: Sure. I think a couple of
10 examples that were already mentioned of Apple and
11 DuckDuckGo as firms that are trying to market privacy
12 as a feature have been raised. Obviously, there are
13 very few. But those are significant examples.
14 In terms of mergers and privacy, I mentioned
15 earlier that data does make merger review slightly
16 more favorable because firms are competing on more
17 fronts. So provided there are at least two firms
18 remaining in the market after a merger and data is a
19 component on which they can use to compete with,
20 competition could still be intense because of all the
21 segmentation that can be done and competition over
22 those segments.
23 MR. SANDFORD: Okay. Does anyone else want
24 to opine yes or no, do firms compete in privacy?
25 (No response.)

1 MR. SANDFORD: Okay. So the other potential
2 antitrust issue I might think of with privacy is -- or
3 privacy and data are, do data serve as a barrier to
4 entry? And is that barrier to entry somehow different
5 than just like I own a factory and you do not, so you
6 have a barrier to entering my industry.
7 So I have a quote here to Darren Tucker and
8 Hill Wellford that states that data are ubiquitous,
9 low-cost and widely available and that an entrant that
10 needs personal data can collect relevant information
11 from its users once a service is operational. Data
12 collected in this manner is free or nearly so.
13 So the argument is sometimes made that, hey,
14 these firms, like these big tech platforms, have lots
15 and lots of data and that makes it harder to compete
16 with them, that might affect competition in some way.
17 A possible counter to that is you can just go out and
18 buy data, you know. There are lots of places you can
19 go buy data. Firms do buy data on where people live,
20 what their income is, how many people are in their
21 household, maybe some information on what their
22 preferences are. And so is there any sense in which
23 data could be a barrier to entry, in which data that I
24 have, but you do not, is irreproducible and gives me
25 an advantage that you do not have?

1 Florian?
2 MR. ZETTELMEYER: So I really disagree with
3 that view. I think that data, in particular back to
4 this discussion of predictions and machine learning
5 and AI, is extremely important. I think what most
6 people do not realize is that the amount of examples
7 that go into being able to train these algorithms is
8 absolutely astronomical. In particular, because in
9 many domains, whether the algorithms get widespread
10 use is very much a function of whether they manage to
11 do predictions in extraordinary ways.
12 In other words, you know, getting an
13 algorithm for predicting correctly 80 to 90 percent of
14 the time may not be a big deal. But if you are at 98
15 percent and you get it to predict correctly 99.9
16 percent, suddenly you have something that is
17 completely usable and creates an enormous change in
18 the way that you can then think of firm strategy of
19 what you compete on, all the services that you
20 produce, it could change the business model that you
21 use.
22 You know, there is this wonderful example
23 that a book that -- a very nice book that recently
24 came out from Avi Goldfarb, who is going to be here on
25 the panel, and Josh Gans' book on what he called --

1 they called "prediction machines," which I recommend
2 everybody to read. In there, they have this very nice
3 hypothetical example of where they talk about the fact
4 that, at the moment, Amazon has, like, a first shop
5 and then ship model. If you could predict to great
6 accuracy what people are going to buy, you could ship
7 first and then shop. That has an enormous effect on
8 strategy, on how you would operate as a company.
9 So I think that those advances are only
10 possible with absolutely huge amounts of data. So I
11 think it is true that more and more data, at some
12 point, has sort of slightly fewer returns, but what
13 you can accomplish with the predictions that arise
14 from that data could potentially be a sea change. So
15 the returns to that additional data is huge. And as a
16 result of this, I think that data is very, very
17 important and it is certainly not ubiquitous in this
18 sense.
19 And we have seen this, by the way, in the
20 search engine wars from a number of years ago, how
21 hard it was for people like Bing to catch up or
22 compete adequately with Google, simply based on the
23 volume of data that they had in order to improve their
24 searches.
25 MR. SANDFORD: Ginger and Alessandro both

1 wanted to weigh in.
2 MS. JIN: Just to play devil's advocate
3 here, we have seen entrants disruptively take over the
4 incumbent although the entrant does not have a data
5 advantage. So we think about Google against Yahoo or
6 Facebook against MySpace. But Florian could be right
7 that maybe, at that moment, that data was not used
8 very efficiently or the data scale had not been large
9 enough and granular enough to have sort of the effect
10 that we observe today.
11 But let's just say, okay, that data is very
12 important today. It is a very valuable asset. It
13 does give an advantage for the incumbent to use that
14 data in a way that has a competitive edge. Okay?
15 Let's say that is true. I think we still need to
16 think hard of how to translate that into, say,
17 antitrust action.
18 Because you can say, okay, in the oil
19 refinery industry you need a lot of investment to
20 start and that means we need to break up the oil
21 companies. I think there is a leap of logic there
22 when you say sort of the barrier to entry is very
23 high. Whether it is in physical assets or in data
24 assets, there is a question we have to ask about the
25 investment that firms are putting into these kind of

1 algorithms or data collections, and they cost money,
2 they cost efforts, they cost talents. And to what
3 extent that is -- we should think that all that should
4 be available to everybody and how would that undermine
5 the investment incentive for the firms to really
6 improve the algorithms and improve the data
7 collection, I think that is a hard question.
8 MR. ZETTELMEYER: It is a very hard
9 question. I think it is also very context-specific.
10 I mean, I do not think this applies to every single
11 context, but I think there are contexts in which, you
12 know, going from huge to extra huge does make a
13 difference. I think it is hard to preview at this
14 time, frankly, when that is the case and when it is
15 not.
16 MR. ACQUISTI: I am not making an antitrust
17 argument because that really is not my field of
18 research or expertise. But it is interesting, I was
19 going in the same direction Ginger was going thinking
20 about examples such as MySpace or Oracle or Yahoo who,
21 notwithstanding having, to some extent, first mover
22 advantage were then replaced by companies like
23 Facebook and Google, and I was thinking what are the
24 differences? To me, there are many. There are many,
25 okay? So it would be simple if it were just one. But

1 an important one is that both Google and Facebook
2 succeeded in creating these two-sided platforms and
3 benefit from network effects on both sides of the
4 platform.
5 If you are an advertiser, you want to be on
6 the platform that offers you great access to
7 publishers. If you are a publisher, you want great
8 access to advertisers. These dynamics are to be self-
9 reinforcing and they create these very, very strong
10 concentration of power in firms, such as Google and
11 Facebook, which may create this potential issue of
12 antitrust, although I am not getting into the issue of
13 then whether it should be split up or so because that
14 is not my area of expertise.
15 MR. WAGMAN: I would also add to that that
16 there are examples of firms scooping up other firms
17 that seemingly have different data, for example,
18 Facebook acquiring WhatsApp for 20-some billion
19 dollars. The data seems different. It seems like a
20 different kind of network. And, yet, the data is
21 extremely valuable. It contains, you know, context
22 lists that can be connected with information Facebook
23 already has about users to better pinpoint users, to
24 better identify them.
25 So this adding up of seemingly disparate

1 graphs or networks or data sets can be extremely
2 beneficial and kind of bring you to that huge point
3 where you can identify people with pinpoint accuracy.
4 MR. ZETTELMEYER: I should also point out
5 that inside the industry, there is actually a concern
6 about this. I mean, there is this open AI initiative
7 that Elon Musk is involved in, which is precisely
8 about trying to make sure that a lot of the advances
9 in that area are in the public domain somehow in order
10 to be able to be shared across everybody because of
11 the fact that there is a concern that you might get
12 too much of an advantage otherwise.
13 MR. ACQUISTI: And to Liad's point about
14 WhatsApp was really great and interesting because it
15 also connects, in a way, to a question that I feel bad
16 we did not fully address, the question about
17 competition. We did not have much to say. But the
18 example of WhatsApp and Instagram is quite interesting
19 from a competitive perspective.
20 Some users started using Facebook less or
21 even migrating away from Facebook to other platforms,
22 such as Instagram also partly, not only, for privacy
23 reasons. And, yet, a powerful company can use the
24 revenues to acquire its competitors -- its more
25 privacy-friendly competitors and reincorporate the

1 data of these users back into their databases. This
2 is an interesting tale about the challenges of
3 privacy-based competition in this market.
4 MR. SANDFORD: Okay. I want to read a
5 quote. So I want to read from a blog post by the CEO
6 of DuckDuckGo, Gabriel Weinberg. The quote, "It is
7 actually a big myth that search engines need to track
8 your personal search history to make money or deliver
9 quality search results. Almost all of the money
10 search engines make, including Google, is based on the
11 keywords you type in without knowing anything about
12 you, including your search history. The fact is these
13 companies would still be wildly profitable if, for
14 example, they dropped all of these hidden trackers
15 across the web and limited the amount of data they
16 keep only to what is most necessary."
17 Okay, this is -- I'm guessing Florian is
18 going to say that is not true based on the data he
19 studied. But this sort of raises the question, is he
20 right? I mean, could we drastically scale back the
21 data, say, Google is collecting from us, just sell ads
22 based on keywords and make a little bit less money,
23 but maybe not that much less and maybe we would be
24 better off by having more privacy?
25 As to the question of the value of targeting

1 ads, I mean, Liad had a -- sorry, Alessandro, in his
2 opening remarks, said that the value of a targeted ad
3 raised revenue by .0008 dollars if I have that right,
4 or maybe there is an extra zero in there.
5 MR. ACQUISTI: There are four zeros.
6 MR. SANDFORD: Okay, one extra zero. So, I
7 mean, there is a question of targeted ads raise more
8 revenue, but how much more? And it sort of seemed
9 like Alessandro is saying, by not very much at all,
10 but Florian's Facebook paper is suggesting that maybe
11 the value is quite substantial. So how should I think
12 about the value of ad targeting? Is it big or is it
13 small and what do we think of the DuckDuckGo guy, who
14 obviously is not an unbiased observer? What do we
15 think of his remarks?
16 MR. ACQUISTI: I am actually curious about
17 what Florian would say about this. I will only
18 comment that the results I was reporting and those
19 found by Florian, they are not contradictory. In
20 fact, they may be very much on the same page. We are
21 looking at what -- at the end of the value chain
22 remains in the hands of publishers. And Florian was,
23 if I understood correctly, looking at how merchants
24 who use certain techniques for advertising can see
25 confluent conversions expand in the presence of

1 targeting.
2 MR. ZETTELMEYER: So I do not know what -- I
3 think my first approach would be to say that Google is
4 in two ad businesses, one is the keyword search ads --
5 keyword-based search ads, and the other are display
6 ads and the display ad networks that they run. So
7 those are different from each other. I believe that
8 while it is true that you may only need keywords in
9 order to place search ads, you certainly need
10 information about users in order to participate in the
11 ad networks and display advertising.
12 So I think maybe that is a little bit lost
13 in that quote. So I do not have, off the top of my
14 head, what percentage of revenue profits in Google
15 depends on one versus the other type of advertising.
16 So I cannot say whether that is correct that, you
17 know, they would still make loads of money if you shut
18 one of the things down or not.
19 By my sense is that in order to do the
20 targeting, you do do this, and I think the big problem
21 is there would be -- I am just a little concerned to
22 the degree that -- you know, Alessandro, I do not know
23 how generalizable this result is about the benefits of
24 targeting. It is just very difficult to get good
25 measurements in this space, I think even for those who

1 are involved in it. I think a lot of times the firms
2 themselves that target do not know how valuable the
3 targeting is.
4 That would be certainly a wonderful area for
5 more research because I do not think we have a really
6 great fact base, frankly, to answer -- to question the
7 gentleman or to kind of challenge the statement the
8 gentleman is posing at the moment.
9 MR. ACQUISTI: I agree.
10 MR. SANDFORD: Ginger?
11 MS. JIN: Yeah, I wonder if the observation
12 you quoted will be related to Florian's earlier
13 comment about this huge versus extra huge. I mean,
14 maybe today, we do not see the extra huge effect yet,
15 but who knows. In the future, there will be
16 technology that can much better use the individual
17 identifiable information from Google versus DuckDuckGo
18 and have a huge lift. I mean, we just do not know.
19 MR. WAGMAN: I would say that from the
20 perspective of economic theory, there is obviously
21 value in knowing more about a consumer. So I could
22 see a consumer, you know, searching for a computer and
23 I know they are predisposed to maybe buying a
24 computer, and then I could maybe know who the consumer
25 is, how much income they have, how much education they

1 have, where they live, whether they have a computer
2 right now or not, and I could use that information to
3 send them to a very different place. Just like firms
4 might steer Mac users to a different list of hotels
5 than PC users.
6 MR. SANDFORD: Okay. I have a couple
7 questions from the audience I will get to. This one I
8 will direct to florian. Florian, if businesses have
9 no good means to evaluate the impact of their targeted
10 ads, why are they spending so much on such ads?
11 MR. ZETTELMEYER: That is a wonderful
12 question. I think that there is, in my experience,
13 enormous amounts of information asymmetry as to -- I
14 think a lot of firms or the people in charge of
15 placing ads in many of these firms are not well aware
16 of this problem.
17 The measurement problem with digital
18 advertising is very pervasive, it is very big. There
19 are a bunch of people who, in academia, have done some
20 amazing work on this, like David Reilly and Garrett
21 Johnson, who is coming tomorrow, and Randall Lewis, et
22 cetera. And you now have an increasing set of people
23 who are very, very sophisticated about thinking about
24 advertising placements and marketing place in general,
25 but the basic problem that exists is that, you know,

1 marketing is a special form of hell when it comes to
2 measurement because of the fact that so much of
3 consumer behavior is highly endogenous and so much of
4 the way that firms target is so endogenous. So
5 measurement, in general, is a very difficult thing.
6 We used to have an area in marketing that
7 was very well measured, which was the direct mail
8 industry. But somehow the people who went into the
9 digital world are not the old mail order guys. Often,
10 they came out of the advertising industry, which did
11 not have as strong a tradition of very good
12 measurement. So there is just a lot of lack of
13 information.
14 I would maintain that part of the problem is
15 that there is a little bit of political economy here,
16 as well, which is that beyond the situation where it
17 is not always clear to me that everybody wants to
18 actually know the answer to how well the advertising
19 is actually working. And I will just leave it at
20 that.
21 MR. SANDFORD: Alessandro?
22 MR. ACQUISTI: Adding a comment to what
23 Florian so eloquently put out and said. Large
24 companies have troubles in understanding the value of
25 targeted advertising for them. Famously, they see,

1 oh, Unilever made some controversial statements about
2 the benefit of social media advertising to them. And
3 these are large companies with very sophisticated
4 research teams. Think about the challenges for medium
5 and even more so small companies that may not have the
6 know-how and skill set available to run the kind of
7 experiments that Florian has been able to run and the
8 larger companies are running to understand the value
9 that they get from that.
10 It goes back to the point that we have been
11 discussing. It is kind of like a red line connecting
12 our different comments of this opacity in the very
13 proposition of certain aspects of targeted
14 advertising.
15 MR. ZETTELMEYER: If I could say one more
16 thing about this, Jeremy --
17 MR. SANDFORD: Mm-hmm.
18 MR. ZETTELMEYER: -- which is that I think
19 what is tricky is that a lot of the advances that have
20 been done with analytics and quantitative methods and
21 machine learning, et cetera, they are advances of
22 prediction. The problem is that -- and predictions
23 work incredibly well in many domains. The big
24 problem, however, is that nearly all marketing
25 expenditure is not a traditional prediction problem

1 because it is a problem of causal inference
2 essentially. In other words, what you want to know is
3 what would have happened had I not placed an ad.
4 And this often does not lend itself very
5 well to sort of organically arising data sets. A lot
6 of people do not understand, in practice, the
7 difference between the fact that something is
8 successful in the sense that it creates a lot of
9 clicks and the idea that what you are really looking
10 for is not whether it creates clicks but whether it
11 creates more clicks than what would have happened had
12 you not done whatever you did. So this deep
13 understanding of causality is surprisingly lacking in
14 a lot of mid to upper-level management areas.
15 I will make this comment later in the panel
16 on the business side a little bit. But it is a little
17 bit as if we have been given the tools to do great
18 data work and now it means that the people who are
19 directing and engaging in using data like this sort of
20 are lacking a little bit of the training to know how
21 to do great data work. So the importance -- this will
22 be my argument later -- the importance of training
23 sort of the decision-making and managerial class up on
24 how to use quantitative methods in order to derive
25 evidence is really important and it is not

1 sufficiently developed at the moment.
2 MR. SANDFORD: Okay. Another question from
3 the audience for Liad and all panelists. Liad, your
4 presentation highlighted the differences in mortgage
5 offerings and opt-in and opt-out locales. We know
6 that there are racial disparities in mortgage
7 offerings across the U.S. To what extent might opt-in
8 or opt-out affect racial discriminatory offerings and
9 to what extent can or should noneconomic variables,
10 like reducing racial discrimination, be factored into
11 these types of data-sharing decisions?
12 MR. WAGMAN: So the analysis did control for
13 race composition. It was done at the census tract
14 level and at the individual loan level. And we did
15 notice the other kind of discrimination in this
16 analysis. For example, certain populations were more
17 likely to be denied a mortgage than others. Now,
18 having said that, the opt-out regime, meaning that by
19 default your information would be traded, had less
20 denials for all groups. Okay?
21 So if we looked at it that way, you know,
22 there are certain benefits that opt-out has that from
23 that perspective. Now, of course, it is kind of --
24 less denials can be looked at as a good thing, it
25 could be looked at as a bad thing because maybe you

1 are matching loans with borrowers in a less efficient
2 way that could cause downstream foreclosures. So
3 there are all sorts of tradeoffs here and racial
4 discrimination is just one of them. It is just
5 another factor and we did control for it in the
6 analysis.
7 MR. SANDFORD: Okay, another audience
8 question. I think I will address this to Ginger since
9 she was the Director of the Bureau of Economics and it
10 is a policy question. Ginger, both in terms of theory
11 and practice, how would you compare ex-post punishment
12 following data breaches versus ex-ante regulation of
13 data practices to minimize breaches?
14 MS. JIN: Very good question. I think there
15 are pros and cons in both approaches. I think ex-post
16 enforcement would give some flexibility for the market
17 to try out new practices and then the Government would
18 not come in until we see a harm to that practice. On
19 the contrary, I think the ex-ante regulatory approach
20 will be really hands-on prescriptive. That is like
21 the Government knows what is going to go on in the
22 near future and you have to do ABC in order to pass
23 whatever threshold I am setting. I think that gives a
24 lot of confidence to the government agency and the
25 employees there to decide exactly what is the right

1 level and how would you define the procedure to reach
2 that.
3 I do think that tradeoff between ex-post
4 enforcement and ex-ante regulation is a very important
5 one and should have much a wider debate among
6 different disciplines.
7 MR. SANDFORD: Okay. Next question. So,
8 you know, if I am an optimist about privacy and sort
9 of big tech companies, I might say something like
10 this. You know, there is a lot more data being
11 collected on me now than there used to be, but it is
12 mostly by companies who give me a product I like for
13 free and the way that they exploit that data is mainly
14 by targeting ads to me. And I do not care that much
15 about targeted ads. It may even be a positive. I get
16 things I am interested instead of random stuff.
17 I think pessimistic scenarios might involve,
18 like, excessive government surveillance or something
19 like that, but there are curbs about that. If I think
20 about big tech companies kind of gobbling up the
21 economy, well, I think, you know, as Ginger mentioned
22 earlier, that companies like Friendster and MySpace
23 and Yahoo and AOL used to be dominant and now they are
24 not. Upstart competitors are able to come along very
25 quickly and with better products and push them out of

1 the market. So I am not that worried about big
2 companies like Google or Facebook because there is
3 competition out there even if there is no company now
4 as big as Google or Facebook. So that is sort of an
5 optimistic view of tech and privacy.
6 What does that view miss, if anything, and,
7 you know, what pushback would you like to give that
8 view, if anything? Liad?
9 MR. WAGMAN: I would say that some products
10 can be made better with data. So for example, if I am
11 on a social network and I see my friends there first,
12 even if we are not connected yesterday, that could be
13 perceived as helpful. In the era of Friendster and
14 MySpace, I do not think data was yet used as part of
15 the product, as part of improving the product quality.
16 Today, it definitely is being used to improve product
17 quality.
18 So entry in this environment is a little bit
19 harder because anything an entrant makes, an incumbent
20 can make as well and use data to make it better. So
21 in that sense, things have changed.
22 MR. SANDFORD: Alessandro?
23 MR. ACQUISTI: I feel that both the
24 optimistic and pessimistic scenarios are both
25 plausible. But I also feel that, going back to

1 something I mentioned at the start of my remarks, we
2 really do not need to choose between the value of
3 analytics and the protection of privacy. We do have
4 tools that go in the direction of trying to achieve
5 both.
6 Once again, I am trying to use language
7 carefully by saying going the direction of trying to
8 achieve both because when you talk about privacy in
9 nascent technologies, you do have to admit they are
10 still young, that they raise some costs. Every time
11 you degrade quality or granularity of the data, you
12 also lower the utility of the data. The interesting,
13 once again, research question for all of us is, if we
14 do use these technologies and they lower the quality
15 of the data and, therefore, they imply some costs, who
16 is going to bear that cost?
17 Is it the consumer through not so
18 well-targeted offers? Is it publishers that run out
19 of business because they cannot sell as targeted ads?
20 Is it merchants that cannot target it as well? Is it
21 data intermediary? Is it society as a whole? Once
22 again, I believe that we do not have yet good answers
23 to these questions and this is where we should put
24 lots of attention on.
25 MR. SANDFORD: Okay. Ginger?

1 MS. JIN: I think one thing sort of really
2 amazing in this space is kind of idiosyncrasy
3 preference. This is not just to say, okay, we all
4 want a safe drug versus a nonsafe drug. It is amazing
5 that different people may have different preferences.
6 Some may be optimistic, some may be pessimistic. Some
7 may sort of have a strong feeling about sort of not
8 giving away my data, but other people would be exactly
9 the opposite.
10 I think the challenge is how can you design
11 a framework to accommodate that kind of heterogeneity
12 but still kind of achieve protection for those who
13 care about it, but also innovations for those that
14 care more about the products coming out of the data-
15 intensive practice.
16 MR. SANDFORD: Okay. This may be a factual
17 question and that is dangerous because you may not
18 know the answer. But going back to the issue of
19 competition between firms and privacy, my factual
20 question is, do firms compete in data security? And
21 the reason I ask that is data security is kind of
22 objectively measurable.
23 I can look at the hash function you are
24 using for your passwords and tell if yours is better
25 or worse than someone else's. It is objective,

1 whereas privacy policies are, one, hard to quantify,
2 hard to measure in any way and, two, if you offer --
3 if your website offers a different set of services
4 than mine does, of course, our privacy policies are
5 going to be different to some extent. So it is really
6 fuzzy to compare my privacy policy to yours, okay?
7 But data security, for example, how you encrypt the
8 passwords that are stored on your server, is
9 objectively measurable. There are hashing algorithms
10 that are better than other hashing algorithms, yet
11 both are used in the market.
12 And it seems to me that I have never seen
13 firms make the claim that we have better data security
14 -- well, okay, never is too strong a word. I do not
15 see firms advertising that I have a better hashing
16 function than this guy so you should come to my
17 website. So, again, it is a factual question. Do
18 firms compete in data security?
19 Florian?
20 MR. ZETTELMEYER: I think it does exist in
21 the B2B space, not in the B2C space as much. So I
22 think if you think about some of the cloud services,
23 like Box and Dropbox, et cetera, they definitely sell
24 themselves as having superior security features and
25 compete on that.

1 MR. ACQUISTI: I agree. There is also
2 potential evidence of some effect in the B2C market in
3 regard to data breach disclosure lows. Sasha
4 Romanosky, who is now with RAND, worked with Rahul
5 Telang and myself on a study on the relationship
6 between data breach disclosure lows and changes in
7 identity theft rates in the United States, across all
8 the states. And there was, indeed, a small, but
9 significant decrease in identify theft.
10 The main variable did not seem to be that
11 the disclosure allows people to actually take action
12 because, as we know, very few people actually take
13 action after receiving a notification in the mail
14 about their records being compromised. But companies,
15 in order to avoid the significant fees associated with
16 disclosure ex-ante, are investing more in security to
17 avoid the data breaches.
18 MR. SANDFORD: Okay. So this is really
19 interesting, the point about B2B versus B2C to me. I
20 mean, in fact, when we do merger review at the
21 agencies, we spend, I would say, the majority of my
22 time since I have come here has been spent on talking
23 to businesses as customers of merging parties, say.
24 So it is interesting to me that B2B customers have
25 strong preferences for data security, but, you know,

1 end user customers like myself might not.
2 Does that suggest that if we think about
3 where antitrust enforcement may need to do something
4 different than it is doing now about data and privacy,
5 would that suggest that it would be mergers that
6 involve businesses as companies? We have one minute
7 left. So that is a good wrap-up question, I guess.
8 MS. JIN: I think there is an information
9 problem similar to what we have discussed before,
10 maybe this is less in the B2B world. If I claim that
11 my cloud has the best security in the whole world and
12 a business customer may, to some extent, confirm that
13 if they have a sophisticated technician to
14 double-check that, but it is almost impossible for
15 individual consumers to double-check that. If we sort
16 of lack that kind of information look-back, then the
17 firms can all claim that we have the best security and
18 then sort of shirk on that claim.
19 MR. SANDFORD: Okay. Would anyone like to
20 avail themselves of the remaining 31 seconds?
21 (No response.)
22 MR. SANDFORD: All right. Then we will wrap
23 up 27 seconds early. Please join me in thanking the
24 panel. Great job, panel.
25 (Applause.)

1 MR. SANDFORD: We will take a -- I believe
2 it is a one-hour break and come back at 1:00 p.m.
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

1 THE BUSINESS OF BIG DATA
2 MR. COOPER: Welcome back from lunch. I am
3 James Cooper. I am with the Bureau of Consumer
4 Protection at the Federal Trade Commission. I will be
5 moderating this panel on the business of big data.
6 So this morning, we heard a lot about some great
7 research in the economics of big data. And so we are
8 going to kick off this afternoon talking about how big
9 data is actually used in a variety of market segments.
10 So we have a great panel to go over this
11 today. We have Christopher Boone, second to my left.
12 He is the Vice President of Real World Data and
13 Analytics for Pfizer. Liz Heier, right next to him,
14 is the Garmin's Director of Global Data Privacy.
15 Marianela Lopez-Galdos is the Director of Competition
16 and Regulatory Policy for the Computer and
17 Communications Industry Association, right next to
18 Liz. Mark MacCarthy, further down there, is the
19 Senior Vice President for Public Policy at the
20 Software and Information Industry Association.
21 Morgan Reed is -- three minutes, two minutes
22 ago, you were not there, I just realized that. Morgan
23 Reed, I have not seen him. So Morgan Reed is the
24 President of ACT, The App Association and he also
25 serves as the Executive Director of the organization's

1 Connected Health Initiative. Next to Morgan is Andrew
2 Reiskind. He is the Senior Vice President for Data
3 Policy for Mastercard Worldwide.
4 And then, finally, to my immediate left --
5 and he is right here because he is going to go first -
6 - is Florian Zettelmeyer. He is the Nancy L. Ertle
7 Professor of Marketing at the Kellogg School of
8 Management at Northwestern University. You have
9 already heard from Florian this morning.
10 So the way this panel is going to work is we
11 are going to -- each of the panelists has between
12 seven and ten minutes, which will be enforced very,
13 very vigorously. And after that, we will hopefully
14 have a vibrant discussion and we will also be
15 collecting as we did in the morning, collecting
16 questions from the audience as we go.
17 So without any further delay, let me hand it
18 over to Florian.
19 MR. ZETTELMEYER: Well, thank you again for
20 having me.
21 So what I want to talk to you about today is
22 not data, per se, but I think a core complementary
23 asset to data, which is the ability as a firm to
24 manipulate it and to use it. And so what I want to
25 start with is first the observation that I am going to

1 call that complementary asset to actually operate and
2 use data analytics -- the terms are getting slightly
3 muddled. Some people are now interchangeably using AI
4 to mean at least a subset of this. But I am going to
5 call it analytics.
6 So the first thing to realize is that
7 basically everybody today has pockets of analytics.
8 There are areas where, for example, the airlines have
9 forever had pockets of analytics and revenue
10 management because this was so crucial for their
11 ability of doing business. The oil companies have had
12 pockets of analytics in oil exploration and assessment
13 of geologic formations, et cetera. So everybody
14 really has them.
15 The trick really is not that they do not
16 exist; the problem is how do you connect them and how
17 do you scale them up at the enterprise level? And
18 that is what a lot of CEOs are worried about is, how
19 do I take this expertise and organize in a way that
20 actually allows us to leverage analytics and,
21 therefore, data at scale?
22 So the point that I want to make today is
23 very simple, which is that I think that companies
24 today are held back by a lack of data science skills
25 at the leadership level. And it is not by the lack of

1 data scientists, that may also be a constraint, but it
2 is a lack of data science skills at the leadership
3 level itself.
4 So in order to make this point, I am just
5 going to start off with an anecdote that I would like
6 to share and it goes like this. So a little while ago
7 I was invited to a thought leadership retreat in a
8 company that operates in the automotive space. This
9 is a company that is partially responsible for placing
10 ads, and as a result of this, has good visibility on
11 how or what consumers do on the online level. And so
12 I was at the car dealership retreat and I had a senior
13 executive of the company who comes up and basically
14 tells that they are excited because they have been
15 able to do something that nobody has been able to do
16 before, which is to link online ad exposure with
17 offline sales, which is a hard thing to do.
18 So this executive comes up and says, let me
19 show you what we found. We ended up classifying
20 people who used search engine advertising into four
21 buckets: People who saw no ads for cars, people who
22 saw dealer ads only, people who saw manufacturer ads
23 only, and people who saw both kinds of ads.
24 And then the exec says, what you see here is
25 the sales conversion rate, the probability that

1 somebody purchases a vehicle after having been exposed
2 to either no ads or dealer ads or manufacturer ads or
3 both ads, and this person says what you can see
4 clearly from here is that the conversion probability
5 goes from 0.7 to 3 to 5, to 14 percent. So this is
6 clear evidence, this person says, that search engine
7 advertising really works and that, in addition to
8 that, it is clear evidence to the fact that dealer and
9 manufacturer ads are complements and not substitutes
10 because 14 percent is more than the sum of 5 plus 3
11 percent.
12 So at this point, there is like an excited
13 discussion in the room, people talk for 15 minutes
14 about what this means for industry and how this can be
15 monetized, et cetera. And then there is a person who
16 says, we should put a press release out about this
17 because this is really cool and nobody has seen this
18 so far in the industry. And so at this point, it kind
19 of goes on for 15 minutes and somebody raises their
20 hand in the room and says, let me ask you a question.
21 Why would somebody not see any correlated ads when
22 they are on a search engine like Google? And the
23 answer of course is, they did not search for a car.
24 And then this person says, so why would
25 somebody see both an ad from a dealer and manufacturer

1 on Google? And the answer of course is, they probably
2 typed in a car name like “Chevy Silverado 1500" and
3 maybe a location that would trigger a deal keyword,
4 like Washington, D.C. And so then this person says,
5 so you are telling me what we have shown -- and he
6 points towards -- you cannot see this here from my --
7 okay, you see this now.
8 But he points to this row here, the no ads
9 column, and says, so tell me what we have shown is
10 that if you are not interested in buying a car, you do
11 not buy a car and pointing towards the very right; if
12 you are really interested in buying a car, you buy a
13 car.
14 So the point about this chart is the
15 following, which is that this data is utterly
16 uninformative about whether advertising works, at all.
17 And the reason is that I do not know whether the
18 difference between 0.7 and 14 percent is driven by the
19 fact that, you know, the people who are on the retail
20 and the manufacturer side and are getting exposed to
21 ads and the people on the left did not or whether it
22 is driven by the fact that they were more interested
23 in buying cars in the first place. Those two things
24 are undistinguishable in this data set. In fact, it
25 is extraordinarily difficult, if not impossible, from

1 this data to say how well does search engine
2 advertising work.
3 And the reason I am bringing this up is
4 because it took the executives in that room 15 minutes
5 and a prompt to realize this was useless data and it
6 should have taken them ten seconds. And if you are
7 trained in causal inference, if you are trained, for
8 example, as an economist or as a social scientist, you
9 see this in ten seconds and start laughing about it.
10 And this is essentially the problem that I
11 am talking about. I have done this with hundreds of
12 executives and it is the norm that people fall for
13 this inference at the beginning without thinking about
14 it more carefully.
15 Okay. What is underlying here is that
16 analytics, the typical view of analytics is that
17 analytics is a big data and a technology problem. In
18 other words, that it is something where you, in order
19 to solve it, you need to invest in big data analytics
20 and technology infrastructure, like Hadoop and Hive
21 and R and Python and whatever; that you have to invest
22 in cloud computing, like, you know, Amazon Web
23 Services or whatever other company is doing cloud
24 services, that you have to invest in data scientists.
25 And I am not saying these things are not

1 important. In fact, they are essential. But the
2 point is they are nowhere close to enough because, at
3 the end of the day, analytics in practice turns out to
4 be mostly a leadership issue. It has to do with
5 things like managerial judgment which there is nothing
6 wrong with the data I showed you. But what is wrong
7 is how you interpreted this data and many people get
8 that wrong.
9 Analytics often has the nasty habit of
10 ignoring organizational boundaries. And so, often,
11 data sharing in companies that crosses organizational
12 silos and profit and loss responsibility is very
13 difficult to achieve and it has to be achieved at the
14 top leadership level in order to create those kinds of
15 alignments.
16 Analytics has to be fundamentally
17 problem-driven. It is really difficult to start with
18 a set of data and saying, let me see if I can find
19 something interesting. It virtually never works in
20 practice. But that means that the people who have the
21 problems need to be involved in actually bringing them
22 to bear on analytics issues, and those are decision-
23 makers and executives.
24 And then the last one is that what a lot of
25 people also do not understand at the executive level

1 is that most of the data that is lying around is
2 actually not particularly useful; that a lot of the
3 data that you need, in particular, as you become more
4 and more sophisticated as a company, needs to be
5 planned and acquired and designed as opposed to
6 collected opportunistically in the normal course of
7 business.
8 So we think that this means that leaders
9 need what we call a working knowledge of data science,
10 which means judge what good looks like, identify where
11 analytics adds value, and lead with confidence. And
12 the consequence of this is that this working knowledge
13 allows you to make the big managerial decisions, like
14 what tools to invest in, what data you need, what org
15 structure you need, and what people you need because
16 in order to link the problems you want to work on and
17 the C-Suite priorities, it turns out this working
18 knowledge allows you to make that link.
19 Thank you very much.
20 (Applause.)
21 MR. BOONE: So it is ten, right?
22 MR. COOPER: Yes, seven to ten.
23 MR. BOONE: So seven to ten, all right. Do
24 not start the clock just yet. Wait one second.
25 (Laughter.)

1 MR. BOONE: I want to make sure I reclaim my
2 time like Maxine Waters.
3 Thanks to the members of the Federal Trade
4 Commission and for the opportunity to provide you with
5 commentary on this very important topic. I would be
6 remiss if I did not acknowledge my distinguished group
7 of fellow panelists on the stage with me here today.
8 But I am going to move on with my comments. I have no
9 slides. So we are just going to talk through this.
10 When it comes to the topic of big data, no
11 industry has felt the weight of this magnitude like
12 the healthcare industry. As the U.S. healthcare
13 system swiftly evolves into a more consumer-centric
14 model, there is considerable interest in increasing
15 access to medical care and therapies for patients,
16 demonstrating value of care and therapies to patients,
17 and improving clinical outcomes with patients.
18 Historically, healthcare provider and peer
19 organizations were in the business of providing acute
20 care to patients under a traditional fee-for-service
21 model. However, each has come to recognize and
22 appreciate the need to understand the genetic,
23 behavioral, social, and environmental factors often
24 referred to as the social determinants of health that
25 contribute to delivering positive outcomes and value

1 for patients.
2 This has, in essence, spawned a new era in
3 healthcare delivery, an era of continual delivery
4 where routinely collected data is continuously fed
5 into a system and ensures we have the information to
6 learn from patient experiences and clinical outcomes.
7 In short, I am referring to the establishment of a
8 learning healthcare system that is built on healthcare
9 informatics, big data, and advanced analytics.
10 So the $64,000 question is why now? The
11 ubiquity of digital health technologies has served as
12 a key enabler for providing this level of care while
13 generating massive amounts of healthcare data or big
14 data. Big data in healthcare is a direct result of
15 the technological advancements in the industry,
16 advancements that include the accelerated expansion of
17 electronic health record platforms, rapid adoption of
18 smartphones and wearable technologies, penetration of
19 social media in our daily lives, cost reductions and
20 genome sequencing, and the repurposing of
21 nonconventional data sources, such as consumer, social
22 economic, and environmental data sets, along with the
23 sophisticated data, analytical tools and techniques,
24 have created an environment where data is a valuable
25 asset.

1 In a broader sense big data in healthcare is
2 often referred to as real world data and it holds the
3 potential to significantly increase the efficiency and
4 effectiveness of all process in the development and
5 utilization of medicines from research and development
6 to regulatory decision-making, to pricing and
7 reimbursement decisions, and even clinical practice.
8 Moreover, real world evidence of the output of the
9 analysis of real world data could supplement the
10 evidence generated from randomized clinical trials,
11 which could considerably improve healthcare decision
12 making for all stakeholders.
13 So what exactly is real world data and why
14 all the excitement? Over the years, the terms “real
15 world data” and “real world evidence” have been used
16 mistakenly as synonymous terms. According to the
17 researchers for the U.S. Food and Drug Administration,
18 the FDA, real world data is defined as data relating
19 to patient health status and/or the delivery of
20 healthcare routinely collected from a variety of
21 sources. These sources typically fall into four major
22 grouping, the first being clinical data, which is
23 patient-level data pulled from electronic health
24 records and/or patient registries that describe
25 treatment in the real world.

1 The second category is administrative claims
2 data, which is the data that is primarily used for
3 billing purposes by providers to insurers or other
4 payors. The third category is patient-generated data,
5 which is data that describes the patient's experience
6 and is collected and shared by the patient his or
7 herself. And the last category is the nontraditional
8 health-related data sources, such as your behavioral,
9 your social media, environmental, and/or socioeconomic
10 data.
11 Real world evidence, on the other hand, is
12 defined as clinical evidence regarding the use and
13 potential differences or risks of a medical
14 therapeutic derived from the analysis of real world
15 data. The simplest way to think about it is real
16 world data is any health data not collected in a
17 traditional randomized clinical trial and can also
18 include data from existing secondary sources.
19 The importance of real world data is
20 critical to all stakeholders across the entire
21 healthcare value chain including physicians, payors,
22 regulatory bodies, patients, and, yes, pharmaceutical
23 and medical device manufacturers. Many are familiar
24 with the use of real world data for informing
25 decisions related to patient treatment options,

1 coverage determinations or even policy options, but
2 some may not be as familiar with how pharma companies
3 actually use real world data. Pharma companies are
4 using real world data and real world evidence across
5 the entire product life cycle to identify targets for
6 the development of new therapies, support regulatory
7 submissions, advance disease understanding and
8 clinical guidelines and support outcomes-based
9 reimbursement decisions.
10 Real world data analysis has been identified
11 by various regulatory initiatives, including the 21st
12 Century Cures Act and the Prescription Drug User Fee
13 Act, as useful supplements to randomize clinical
14 trials. Specific applications include the
15 acceleration of drug approval pathways and expanded
16 indications for approved medical therapies. When
17 it comes to the process of collecting and analyzing
18 real world data, generally, we think of it in three
19 stages.
20 The first stage is the study planning, which
21 is where we seek to understand the evidentiary needs
22 of key stakeholder support groups, such as a regulator
23 or a payor. We then formulate a research question
24 that then feeds into a study designed where we
25 identify the appropriate data sources to conduct that

1 study. Now, it is equally important as part of this
2 processing to assess the availability, accessibility,
3 portability, and even quality of the data for that
4 particular study.
5 The last stage is where we actually
6 communicate and socialize the actual results of that
7 particular study through a scientific publication.
8 From the perspective of Pfizer, we primarily connect
9 deidentified data to use in our real world data study
10 analysis from third-party data aggregators. If there
11 are any data linkage and/or aggregation activities
12 required, we work with these aggregators, who possess
13 the technical expertise and competency, to effectively
14 collect, manage, and link the patient data.
15 Now, the benefits of analyzing real word
16 data for consumers or patients generally we feel is
17 tremendous. We live in the world where most of the
18 health-related data is collected outside of the walls
19 of a provider organization. For example, consumers
20 now possess apps on their smartphones that allow them
21 to perform tasks such as recording daily vital signs,
22 documenting daily food intake, and even detecting
23 triggers or symptoms for certain clinical events.
24 These real world data sources and studies that are
25 associated with it are vital to documenting and

1 understanding the benefits and risks of medical
2 therapies in a heterogenous population and to
3 determining whether patients in routine clinical
4 practice are achieving positive outcomes.
5 As is often the case with cutting-edge
6 scientific and technological advancements, a full
7 understanding of the ethical and policy-oriented
8 implications lags behind. There are several key
9 considerations to keep in mind as we think about big
10 data privacy and competition. Quite frankly, I do
11 believe many of the key policy and ethical
12 considerations are pretty much industry-agnostic,
13 which means that we tend to all deal with the same
14 major issues.
15 At the high level, the issues that are well
16 documented are around informed consent and privacy.
17 Some other concerns that are starting to bubble up are
18 issues around data ownership or the rights to use the
19 data, the appropriateness of methods to analyze the
20 data, the appropriateness of the question being
21 analyzed, and even the legal context for which this
22 analysis takes place.
23 According to a 2017 consumer voices survey
24 conducted by Consumer Reports, 70 percent of Americans
25 lack confidence that their personal information is

1 private and secure. Ninety-two percent of Americans
2 think companies should have to get permission before
3 sharing or selling their online data and 92 percent of
4 Americans think companies should be required to give
5 consumers a list of all the data they have collected
6 about them.
7 Privacy concerns related to allowing the
8 access and analysis with large real world data sets
9 have greatly limited its potential. Since pharma
10 manufacturers do not generate real world data
11 directly, data access, data availability, data
12 portability and data quality remain significant
13 barriers to advancing the science.
14 Other ethical considerations that the FTC
15 should keep in mind are the existence of big data
16 divides, which is created between those who have or
17 lack the necessary resources and infrastructure to
18 effectively analyze these large data sets. The next
19 one is the monetization of data and the potential
20 problems with ownership of intellectual property
21 generated from the analysis of these aggregated data
22 sets.
23 And lastly, the future of real world data
24 and evidence is in the aggregation of genomic and
25 other “omic” data and the possible dangers of

1 intentional or unintentional group level ethical
2 harms, specifically as it pertains to patients'
3 beliefs about the benefits or harms to a particular
4 racial or ethnic group in studies.
5 There is considerable high hopes for the use
6 of real world evidence to improve decision-making in
7 the U.S. healthcare system, but all stakeholders have
8 a role to play. Pharma manufacturers have a critical
9 role in driving innovation by using real world
10 evidence to support clinical trial designs and
11 observational studies to generate evidence and new
12 treatment approaches. However, the need to protect
13 personal data, consent, ethics, and data access are
14 equally important and harmonization of public policy
15 and legal frameworks will be necessary to realize the
16 full value of real world evidence.
17 It is critical that the FTC, as part of its
18 role to protect consumers and promote lawful
19 competition, take affirmative steps to promote ethical
20 use, data ownership and privacy as its pertains to big
21 data and healthcare. These are important
22 considerations to keep in mind as the FTC reviews the
23 state of big data in business and how it affects
24 consumer privacy and industry competition. Pfizer
25 stands ready to discuss the shared responsibility with

1 all interested parties to make this vision a reality.
2 Thank you.
3 (Applause.)
4 MS. HEIER: I am a little bit shorter.
5 Well, first, I want to say thank you to James Cooper
6 and the rest of the FTC staff for inviting me to
7 participate today.
8 My name is Liz Heier and I am the Director
9 of Global Data Privacy at Garmin. It is a bit of a
10 coincidence that I am following Chris since we are a
11 wearables company.
12 My 11-year tenure with Garmin did not start
13 in data privacy. My diverse IT experience includes
14 software development, both as an engineer and a
15 manager, incident management, and data security.
16 These roles have given me a unique perspective on the
17 multifaceted issues corporations face in the areas of
18 data protection and privacy.
19 Garmin was founded in the Kansas City area
20 in 1989 by Gary Burrell and Min Kao, whose belief in
21 the potential of using GPS in avionics and in consumer
22 electronics was not shared by their then current
23 employer. They believed so strongly in the product
24 they were creating that they named the company after
25 themselves by combining their first names, Gary and

1 Min. This was long before Hollywood came up with
2 Brangelina and Kimye.
3 (Laughter.)
4 MS. HEIER: Since its founding, Garmin has
5 grown into a global company of over 12,000 employees
6 spread across 60 offices worldwide. We create
7 products in five market segments, aviation, marine,
8 sports and fitness, outdoor recreation, and
9 automotive. We recently shipped our 200 millionth
10 device.
11 Over the last three decades, Garmin has
12 grown and thrived through its innovation, ingenuity
13 and diversified product lineup. In the 2000s, a
14 majority of our revenue came from our automotive
15 personal navigation devices which sat on our
16 consumers' dashboards. By the time that product
17 became saturated and turn-by-turn directions were
18 ubiquitous on mobile phones, Garmin was ready with
19 new, market first products in our other segments.
20 We have seen phenomenal growth in our sports
21 and fitness segment in recent years with the
22 popularity of our wearables and their companion mobile
23 apps, websites, and services. As I mentioned
24 recently, Garmin recently shipped its 200 millionth
25 device. It was only six years ago that we crossed the

1 100 million mark. Much of that rapid increase can be
2 attributed to the popularity of our wearable products.
3 Many of the owners of these wearables
4 choose to provide their data to Garmin through our
5 mobile apps to enhance their user experience. This
6 means that Garmin has been entrusted with the personal
7 data of millions of users from nearly every country
8 in the world. At Garmin, we believe the data that
9 our customers create and upload through our apps and
10 services belong to our customers. We believe that
11 these apps enrich the user experience of our devices
12 and, in turn, enrich the lives of our customers,
13 whether their goal is to become healthier, share
14 their adventures with friends or fans, or travel more
15 safely in the water, in the air, on the road or on the
16 trail.
17 Garmin makes money selling our devices and
18 we have no need to monetizize our customers' personal
19 data to be profitable. It is not in our business
20 model nor our corporate culture to sell customers'
21 personal data. Today's constantly evolving technology
22 allows our devices to record increasingly detailed and
23 powerful data sets. Through the sensors in our
24 wearables, our customers can monitor their heart rate
25 in real time, as well as view graphs of historical

1 values and averages, all of which could reveal
2 indicators of potential medical issues, such as sleep
3 apnea or atrial fibrillation.
4 Our devices can detect a bicycle crash and
5 automatically alert a user's emergency contract with
6 his or her GPS location and our devices can help
7 consumers navigate hostile terrain while sending text
8 messages to their loved ones to let them know all is
9 safe or to call for help if it is not. These are
10 critical services to many of our customers. But the
11 data required to provide them could be harmful if
12 publicized or misused.
13 We recognize that our customers put their
14 trust in Garmin when they share their personal data
15 with us. We believe that our customers should have
16 the ability to make informed choices when deciding
17 when and how much data to share.
18 A large majority of our products can be used
19 fully out of the box without ever connecting to the
20 internet. For those customers who do choose to use
21 our apps and services, all sharing options are set to
22 private by default and many individual features can be
23 turned on or off, thereby putting the customer in
24 control of what personal data are processed.
25 If the customer decides to no longer use our

1 services, he or she could delete their data at any
2 time. We also do not share their data with anyone
3 unless our customers ask us to do so, nor do we
4 constantly track the location of every Garmin device
5 on the planet. So as much as we would like to help
6 your lost or stolen Garmin device, we just cannot.
7 When the GDPR was approved by the European
8 Parliament in 2016, as was true for many companies, it
9 was Garmin's legal team that began to campaign our
10 leadership and our board of directors that the GDPR
11 issue was big, hairy, and not going away. Our
12 leadership got the message and soon realized that data
13 privacy was not only a legal concern, but something
14 that would have to be integrated into our culture.
15 And that is where I came in.
16 I am not a lawyer, I am a software engineer.
17 Who better to work with engineers on the GDPR than one
18 of their own? With a strong governance team of key
19 executives, business leaders, and legal counsel
20 supporting me, we used a risk-based approach to create
21 a compliance program that was guided by pragmatism,
22 transparency, and usability. In that spirit, Garmin
23 supports a federal privacy law that would preempt
24 state law and position U.S.-based businesses to better
25 compete in a global economy.

1 The GDPR is not perfect, but there are many
2 things it gets right, and any U.S. company that does
3 business in Europe has already invested in complying.
4 Garmin alone invested more than 800 person-months of
5 effort to ensure compliance. Consistency and data
6 privacy laws benefit everyone by lowering the cost of
7 implementation, reducing complexity, and allowing for
8 globally recognized and understood paradigms.
9 One of the things I believe GDPR got right
10 was that it largely harmonized data protection
11 regulations across the EU. Prior to the GDPR,
12 companies that do business across Europe had to
13 navigate the complex data protection regulations of
14 all EU member states. This resulted in confusion,
15 inconsistencies among the various regulations, and a
16 higher cost of compliance. Having a harmonized
17 regulation in the EU, even one that sets a very high
18 bar like the GDPR, brings much-needed certainty to all
19 involved, including the regulators, the businesses,
20 and the consumers.
21 Without a federal privacy law in the U.S.,
22 we would risk going backward to a place like the
23 pre-GDPR European Union where companies could be
24 forced to comply with numerous, possibly inconsistent,
25 state privacy laws. We have seen California recently

1 enact a privacy law and the trend will almost surely
2 expand to other states in the absence of a federal
3 privacy statute that preempts state privacy law.
4 A federal privacy law would also pave the
5 way for trusted transfers of data between the U.S. and
6 the EU without the uncertainty of yearly assessments
7 and frequent challenges to available transfer
8 mechanisms, like Privacy Shield and standard
9 contractual clauses. Like Garmin's services, today's
10 economy is global and it is cost-prohibitive for
11 companies to maintain localized data centers for every
12 country. We need trusted and stable methods for data
13 transfer that allow personal data to be stored in and
14 managed from locations where resources, both technical
15 and personnel, are available.
16 In closing, the personal data and associated
17 processing activities, including big data, provide
18 valuable, often life-altering, benefits for our users
19 whether they are taking their first steps towards a
20 healthier lifestyle or are training for next Ironman
21 Triathlon. Adequately securing their data and
22 handling it responsibly and transparently is a duty
23 that we take very seriously. We support federal data
24 privacy legislation that would promote consistency and
25 align with today's global economy.

1 Thank you.
2 (Applause.)
3 MS. LOPEZ-GALDOS: Hi, good afternoon,
4 everyone. My name is Marianela Lopez-Galdos. I am
5 the Director of Competition and Regulatory Policy at
6 the Computer & Communications Industry Association,
7 and we represent big and smaller tech companies from
8 the U.S. and elsewhere. Before I get started, let me
9 thank James Cooper and the FTC for inviting me to be
10 here. It is a great opportunity for us, but also the
11 FTC more broadly for putting together all these
12 hearings. I know there is a lot of effort behind it,
13 so we really commend you for that.
14 So we are trying to understand how companies
15 use data and I think what I am going to try to do here
16 with my brief remarks is try to explain to you the
17 role that data plays for data-driven companies like
18 the ones that operate in the digital economy. And I
19 bring here today with me three ideas.
20 First, that data is not essential, that
21 ideas are. Second, that in the digital economy,
22 innovation rather than market positioning is more
23 important. Finally, that as technology progresses, we
24 will see that the need for data will diminish, so
25 therefore we need to be very careful and ensure to

1 preserve the incentives for companies to keep
2 innovating in this industry.
3 So let me get started with my first idea.
4 What do I mean by saying that data is not essential,
5 that ideas are? What I mean is that similar to the
6 brick-and-mortar world, in the digital economy,
7 companies exist, flourish and compete because they
8 have a good idea and then that idea allows them to
9 bring to the market a product and a service that
10 consumers like. Therefore, it is not access to data,
11 what allows these companies to compete and to exist,
12 but, rather, the initial idea.
13 So we need to clearly understand that an
14 idea comes first. And this idea that I am -- what I
15 am saying about data being essential seems very
16 obvious, but we sometimes forget when we discuss the
17 role of data and the role that data has for the
18 digital economy that successful winners exist not
19 because they have access to data, but, actually,
20 because they bring to the market something, a product
21 or a service, that consumers lacked. And we have many
22 examples of these in the market if we look at recent
23 history. For examples, you can see how Snapchat or
24 Slack basically became very successful companies
25 without having access to data in the beginning. We

1 also see how Handshake has become a very strong
2 competitor to LinkedIn with more than 14 million users
3 right now among recent graduate students.
4 And we will have an opportunity to listen to
5 Catherine Tucker, I think, later this afternoon and we
6 have been listening to her during these hearings, also
7 to Professor Lambrecht, and I think in a paper they
8 published recently they have a quote that I would like
9 to share with you because it really summarizes the
10 idea that I bring with me today for you.
11 The history of the digital economy offers
12 many examples like Airbnb, Uber and Tinder, where a
13 simple insight into consumer needs allowed entry into
14 markets where incumbents already had access to data.
15 So this is how we summarize my idea that data is not
16 essential. But there is something more that I would
17 like to share with you today, which is that the more
18 access to data, it does not bring added value to some
19 companies.
20 So there is -- Stanford University conducted
21 a study with a set of images from dogs. And they
22 managed to prove that more data gives you better
23 results in data analytics, but to a certain extent.
24 There are limited dimension return for companies when
25 analyzing, for example, images. And I am happy to

1 discuss more about the Stanford study later during our
2 discussion.
3 But, you know, if you think about our own
4 personal experiences, imagine when you were trying to
5 buy a car and you spent six months looking into cars
6 in the market or looking into different brands as the
7 first speaker explained today. So that data becomes
8 late as soon as you purchase the car. So the value
9 of data is quite limited. And, therefore, we need
10 to be very careful with those who argue that data
11 is an essential input because that rests on a
12 misunderstanding of the concept of data and the role
13 that data represents at least for the digital economy
14 and data-driven companies.
15 And that leads me to my second idea, which
16 is that innovation rather than market positioning is
17 what drives the digital economy. What do I mean by
18 this? If we accept that data has limited diminishing
19 returns and that it is not essential, then we can
20 actually understand that data cannot be used to drive
21 a competitor out of the market.
22 So how do companies compete with data?
23 Well, what they do is invest in what I want to call
24 today here, data-driven R&D. They really need to
25 invest and understand data analytics. Because once

1 they have access to data, if they do not have the
2 right analytics and the right decision-making
3 processes for the results that data analytics gives
4 you, that data is basically useless. So that is how
5 companies compete, investing in R&D, investing in
6 innovation.
7 And basically that leads me to my third and
8 last idea which is that as technology progresses, we
9 see a lot of advances. We have come from the IBM
10 linear computing to quantum computing and now we are
11 talking more about machine learning and more broadly
12 AI, but what we are really talking about is machine
13 learning. And in machine learning, data analytics is
14 fundamental.
15 If we speak to engineers working in this
16 area, you will learn that they are progressing quite
17 significantly in the last years. And, for example,
18 now, you will hear them talk about synthetic data,
19 where they use kind of artificially-created data that
20 does not track back individuals, so confidentiality
21 and privacy no longer becomes an issue. But, also,
22 you will hear them speak about zero shot learning
23 which is basically a methodology used by machines to
24 recognize objects without having been trained or
25 without having received label training to recognize an

1 object. So, for example, a machine will be able to
2 distinguish a zebra from a horse without having seen a
3 zebra before. So this is what is happening in the
4 digital economy and this is where technologies --
5 digital companies are investing money and they are
6 advancing quite quickly.
7 So if we understand that with the progress
8 of this technology, the access to data will diminish
9 over time -- the importance of access to data will
10 diminish over time, we understand how important it is
11 to preserve the incentives to innovate and how
12 important it is for our progress and for the future of
13 AI and machine learning to make sure that we do not
14 intervene in data-driven markets unless there is
15 actual harm to consumer. And by preserving these
16 incentives to innovate, we will make sure that we can
17 keep progressing for our society. And with this idea,
18 I stop here and I look forward to our discussions.
19 Thank you.
20 (Applause.)
21 MR. MACCARTHY: So my name is Mark
22 MacCarthy. I am with the Software and Information
23 Industry Association. And I want to thank the
24 organizers of this workshop, Jim Cooper and others,
25 for inviting me to be here today to talk about these

1 data analytics issues.
2 I liked the phrase that you used “analytics”
3 rather than AI or machine learning. It covers a
4 broader range of things.
5 Let me tell you a word or two about SIIA.
6 We are a technology trade association. We have three
7 groups of members, one group of the traditional
8 technology companies, companies like Adobe and Intuit,
9 Red Hat, although Red Hat just got bought, I think
10 Google and Facebook, and then we have information
11 service companies, companies like LexisNexis, Thomson
12 Reuters, Refinitiv, which used to be part of Thomson
13 Reuters, Dun & Bradstreet, and we have ed tech
14 companies, companies that provide personalized
15 learning services to schools, and that is companies
16 like Pearson and McGraw-Hill and Cengage.
17 I want to talk to you today a little bit
18 about some of the uses of data and analytics that
19 these companies are involved in, and I want to talk
20 about four specific cases and just remind you at a
21 high level the kinds of things that are being done
22 today with data and analytics.
23 So the first one I want to talk about is the
24 production of fair and more accurate credit scoring
25 models. The second is the increase in speed and

1 effectiveness of student learning caused by
2 personalized learning technology. The third is the
3 improvement in online personalized ads caused by the
4 new machine learning techniques, and the fourth is the
5 improvement in business risk analytics that is taking
6 place today.
7 So first, general remarks. There is a new
8 development in the data analytics world, but it is a
9 natural evolution of the older techniques. There is a
10 lot more data that is available. It is different
11 kinds of data and the speed at which the data becomes
12 available is much more rapid. So the techniques used
13 for processing this data are different. And the key
14 thing is that the new techniques allow the detection
15 of patterns that would not be available to human
16 intuition and that are not based on prior hypotheses
17 that are developed by researchers. They emerge, so to
18 speak, from the data itself. While the results are
19 sometimes startling, it turns out that the policy
20 issues that are raised by these newer data analytics
21 technologies are much the same as the older policy
22 issues.
23 So with that as a general remark, let me get
24 into the discussion of credit scoring. You all are
25 probably familiar with credit scores. The credit-

1 scoring models have been used for generations. They
2 increase the accuracy and fairness of credit-granting
3 decisions, certainly compared to the human judgment of
4 loan officers who often use subjective assessments.
5 But the traditional credit scores have limits. They
6 do not effectively provide scoring for almost 70
7 million Americans because they rely heavily on data
8 that is from credit reports and that relies mostly on
9 payment information. And this deficit adversely
10 affects, historically, disadvantaged minorities. A
11 study by LexisNexis found that 41 percent of that
12 population could not be scored by traditional credit
13 scores.
14 So they developed their own credit-scoring
15 model, largely by going to new sources of information,
16 new data sources, educational history, home ownership,
17 court records. And with this new availability of
18 data, they found that they were able to score fully 81
19 percent of previously unscorable applicants for
20 credit. And this example shows that even just
21 expanding the kind of data being used and not really
22 using dramatically new modes of analysis can
23 dramatically improve outcomes.
24 In the credit-scoring world, there are also
25 machine-learning models that are being developed by

1 researchers and they will soon be ready for deployment
2 in practice.
3 The second area I want to talk about is
4 personalized learning. Researchers have shown that
5 many students who eventually drop out of high school
6 can be identified as early as sixth grade. And the
7 basis for this identification is their behavior, their
8 attendance in classes, and their, of course,
9 performance. Even more can be identified by the time
10 the students reach the middle of ninth grade.
11 Now, early warning indicators based on these
12 data points can be used and can generate risk scores.
13 This knowledge will allow schools and teachers to
14 provide these students at risk more meaningful
15 interventions and support. And when this happens, it
16 increases the number of students that graduate ready
17 for success either in further schooling or in their
18 careers. In one school in 2013, fully one-third of
19 the students who were being flagged for being late at
20 school or missing school got back on track after these
21 remedial programs.
22 Personalized learning also will help target
23 students according to their learning styles and bring
24 to them the best available learning techniques. In a
25 developmental math program, math courses, used in one

1 community in Chicago, a program called ALEKS, which is
2 produced by McGraw-Hill, uses artificial intelligence
3 to help students progress through the material and it
4 adapts the material to their learning needs. The
5 schools that are using this program report that this
6 new technology gets students through their remedial
7 material much more rapidly than traditional methods.
8 So let me move on to the third area,
9 improved personalization for online ads. This really
10 takes place at two levels. One is the analysis of
11 website movements, which can aid websites in providing
12 material, content material, and ads, and improved
13 analyses of large customer databases. Now, we are all
14 familiar with this, the movement of website visitors
15 on a website is usually recorded and it contains data,
16 such as which pages are visited, how long you spend on
17 which page, how you shift from one to another, the
18 sequence and so on, and critical patterns of that kind
19 of usage that cannot be identified by human beings or
20 by eyeball inspection of the data that can be inferred
21 through machine-learning programs.
22 And once these patterns are discovered,
23 website visitors can be segmented into different
24 groups based on the preferences that are inferred
25 about them and the website's content can be

1 personalized to those preferences and the ads that are
2 served to them can be personalized to their interests
3 and needs.
4 A second way, companies often have large
5 aggregations of their own consumer data or they can
6 obtain them readily from third parties, and they need
7 an effective tool that can detect patterns in the data
8 that will enable them to become better at their
9 marketing campaigns. Now, machine-learning programs
10 can dig through data to find insights that can be used
11 to devise smarter and more effective ad campaigns.
12 They are so good that they can also advise marketers
13 what type of campaign to use, whether it is email or
14 social media engagement or online advertising or
15 recommendations on websites.
16 In addition, the use of inferred
17 psychological characteristics is often a good
18 mechanism for improving the effectiveness of
19 advertising. The level of extroversion, for example,
20 or openness can be inferred from social media
21 behavior, and if you match the content of advertising
22 to this characteristic, you can improve responses
23 significantly, according to one study, an increase of
24 40 percent more clicks and up to 50 percent more
25 purchases.

1 Now, of course, the benefits of these
2 increasingly effective target ads is the ease and
3 convenience of consumers who are seeing material that
4 is more appropriate to their need. But, also,
5 additional revenue to provide ads supported free or
6 subsidized content.
7 Let me shift to my last topic, improved
8 business risk management services. Information
9 service companies help their business customers to
10 manage their risks using data sets that they have
11 acquired in various ways. These data sets usually
12 rely on public records and information about people in
13 their business capacity, their status as directors or
14 officers or stockholders of companies, and they also
15 include lots of nonpersonal information, such as the
16 financial and operating characteristics of companies,
17 including how well they have paid back their own
18 debts.
19 Now, the predictive analytics component of
20 this includes the likelihood of repayment of a
21 business loan or a profitability analysis that would
22 assist a company in a merger analysis. These
23 techniques also help companies make better decisions
24 and manage risks, like identity theft, fraud, money
25 laundering, and terrorism. Regulators also want

1 financial institutions to detect terrorist financing
2 and money laundering using whatever techniques are
3 most effective.
4 The coming thing in this area is that the
5 same machine learning-techniques that can spot a
6 pattern of bad transactions in the credit card world
7 can also be used to assess the risk that a potential
8 customer would engage in these kinds of suspicious
9 activities.
10 So that is my quick survey of the areas
11 where big data and analytic techniques are improving
12 things. As I say, one of the major policy take-aways
13 is that while these are new techniques and sometimes
14 produce startling results, I do not think they raise
15 fundamentally new policy issues. And let me put off
16 the discussion of those policy questions for the give
17 and take later on in the discussion.
18 Thank you for listening to me.
19 (Applause.)
20 MR. REED: Well, given the number of people
21 I have been watching slowly move their eyes down to
22 their smartphone, hopefully looking to an app while
23 they are there, I realize that we have started to hear
24 some of the same stories from panelists as we have
25 gone down the line. So I decided to try out some of

1 my notes and kind of try to weave a little bit of a
2 story into what we are up to in the healthcare space,
3 but that maybe can run some threads and even ask some
4 questions for my own panelists for the later session.
5 So my name is Morgan Reed and I am the
6 President of the App Association. And I hope -- well,
7 a quick show of hands so we can throw out the
8 infidels. Anybody here not have a smartphone?
9 Excellent, thank you all for keeping me fully
10 employed, I love you all. It is great to have you
11 here.
12 Here is the thing, the technology that I
13 work on and the industry that I help to lead as the
14 President of the App Association is the fastest
15 growing technology in the history of mankind. Full
16 stop. We have successfully put access to the world's
17 connected information into the fingertips of roughly
18 two billion people and we have done it in less than
19 ten years. It is faster than fire, it is faster than
20 the wheel and faster than the next fastest adopted
21 technology which was the microwave, kind of cool,
22 actually, the microwave, the second most fastest
23 adopted technology after the smartphone.
24 So with all of this access to information
25 and such a life change revolutionary idea, information

1 in the hands of more people, there is a concomitant
2 secondhand to that, which is data. What do those
3 people know? How are they feeding back into this
4 collective system? So when with you all my fellow
5 panelists talk about their kind of segmented chunks of
6 big data and the way that they use it and how well we
7 are protecting it and we are making sure we are being
8 very careful with it, all of that is true. But what
9 we have not really talked about here -- and Mark hit
10 on it and we have touched on it a little bit -- is
11 this is kind of amazing. This is life changing for
12 billions of people in a good way.
13 Yes, we need to protect it. Yes, we need to
14 be careful with our regulation. Yes, we need to think
15 about how it implies and what it implies when it comes
16 to competition. But let's not lose track of the fact
17 that it is life-changing and life-beneficial to
18 billions of people around the world. So let's dig
19 into some specifics.
20 I head up the Connected Health Initiative
21 element that we are part of and let me tell you some
22 really depressive things about America that hopefully
23 will not cause people to start drinking until after it
24 is over, but by 2025, the United States will be 90,000
25 physicians short, 90,000 physicians short. By 2030,

1 we will have 70 million Americans over the age of 65.
2 And I will give you a little secret, people over the
3 age of 65 are sicker.
4 Two weeks ago, I testified before the Senate
5 Health Committee, Senator Enzi is the chair. My
6 colleague to the right was the insurance commissioner
7 from the State of Wyoming, who, great guy, former
8 rodeo rider, said in that drawl, the State of Wyoming
9 currently only has 157 physicians for every 100,000
10 people. If you are sick in the State of Wyoming,
11 leave the state to get care.
12 So the question that we need to be thinking
13 about when we are asking these questions about big
14 data and the business of big data is, what does it
15 provide to people? To consumers? And I am here to
16 tell you that the demographic numbers are clear. If
17 we do not find a way to engage with digital medicine
18 and big data, we cannot support the number of people
19 in this country who will need quality care. Cannot do
20 it. No amount of money, no amount of change will get
21 the number of physicians that we need to have in
22 practice.
23 So a little bit more upbeat, right? We can
24 do stuff with data. So what do we do with it and what
25 are some examples of how we move forward? Well,

1 primarily, I thought one of the interesting things my
2 first panelist said was, well, we only need structured
3 data. Anybody in this room, do not show hands because
4 it is medical information, but I am going to assume
5 that everybody in this room has someone that they know
6 that has some kind of autoimmune disorder, whether or
7 not it is one that is related to chronic fatigue
8 syndrome, rheumatoid arthritis, any of the other
9 concomitant diseases that go along with it, impacts
10 from Hashimoto's thyroiditis, all of those cases we do
11 not actually know what is wrong with you. That is the
12 depressing part.
13 Anybody who has an autoimmune disorder, you
14 go to your rheumatoid arthritis specialist and they
15 say, well, let's try this. And part of the reason why
16 is that physician that you are seeing, so they are at
17 the top of your game -- a physician at the top of
18 their game has seen roughly 29,000 patients by the
19 time that they get to you. Of patients that will have
20 your identical comorbidity, your genetic type, your
21 age, your other key factors, where you live,
22 everything else, you are lucky if your physician has
23 seen 500 people that look like you.
24 So that means your physician is going to
25 base your treatment off of what they learned 15 years

1 ago in school, that continuing education class they
2 took a booze cruise somewhere and, hopefully, 500
3 points of data. You should be angry at that. The
4 fact that everybody on this panel has talked about how
5 we are absorbing and utilizing big data and, yet, why
6 is it that your physician is making a treatment
7 decision based on 500 minuscule points of data that
8 you hope will be relevant to your condition?
9 So as you consider the question of the
10 business of big data, the question that you should be
11 asking is how do we use the business of big data to
12 actually produce a better consumer outcome? And in
13 the healthcare space, I am going to offer a couple of
14 very obvious examples that we are working on right
15 now.
16 Through the Connected Health Initiative, we
17 work with academic medical centers, businesses, the
18 American Medical Association, patient groups and
19 others. One of the leading areas that we have real
20 difficulty in in this country is, of course, type 2
21 diabetes. It is an epidemic. It is one that we know
22 how to solve and, yet, people keep continuing the same
23 behavior.
24 So what can big data provide us in terms of
25 insights? Well, there is a company out of Georgia

1 called Remedi, but spelled with an I. Look at it on
2 Twitter. You will be able to see it after this
3 session. They are actually using remote patient
4 monitoring data from wearables, like yours and others,
5 to paint a picture of a person. And here is where the
6 analytics and the big data comes in and really makes a
7 difference.
8 Through something called clinical division
9 support, they actually allow a physician to model the
10 treatment of the patient before prescribing it to
11 them. They actually take in the data from your
12 electronic health record, combine it with wearables
13 information and they create patterns and they say,
14 well, this treatment schedule has about a 60 percent
15 chance of likelihood of success. This one, we see
16 that a person with these similar conditions, you are
17 likely to see this outcome.
18 The decision is still in the hands of the
19 physician, hence the support part of clinical division
20 support, but ultimately allows that physician to bring
21 in multiple data sets, look at it, overlay it, and
22 instead of going to the doctor and saying, take these
23 pills and three weeks later we will see how you do,
24 they are able to run multiple scenarios prior to your
25 treatment so they get closer to the right answer.

1 Now, that requires large data sets -- and
2 something Christopher said that I am always a little
3 concerned about is this idea that, well, you know you
4 have got real world data, how do you bring it in and
5 integrate it? I think all health data needs to have
6 that real world element in. Because where we live,
7 what we eat, what our genetic situation is, is all
8 part of figuring out how to be healthy.
9 And Chris said something else, he talked
10 about consumers. One of the parts that we -- that
11 Christopher and I know in this case is this difference
12 between patient and it has to do with how you are
13 paid. But I realize that a patient is actually a
14 person. None of us want to be a patient, right? If
15 we are sick, we want to get healthy; if we are
16 healthy, we want to stay healthy. And what we need to
17 look at is how does big data get us there. So a
18 product like Remedi helps to get us there.
19 Earlier on, you said that, you know, we need
20 to structure all that data, but one of the things that
21 we have learned is if we do not know the answer, then
22 I cannot necessarily structure the data the right way
23 to answer the question. But I did agree with that
24 first points, which is it is all about asking the
25 right questions.

1 So as we go through the rest of the panel
2 and go through the Q&A, we will talk a lot about how
3 do we provide short form notice and what kind of
4 consent mechanisms do we need and what are the
5 regulatory necessity of the GDPR or other elements.
6 But the primary question you should be
7 asking is, how does big data actually produce an
8 outcome that is good for consumers/mankind, for
9 patients. Because, right now, the medical care you
10 are getting that does not rely enough on big data
11 should not satisfy you. You need to ask more of your
12 data and ask more of the healthcare system that can
13 use that data because we can do better using big data.
14 Thanks.
15 (Applause.)
16 MR. REISKIND: Good afternoon, everyone.
17 Thank you, Morgan, for waking us all up with good
18 news. It is always good to get good news after lunch
19 and keep everybody awake.
20 So my name is Andrew Reiskind, Senior Vice
21 President for Mastercard. I am responsible for data
22 strategy and innovation. And so who is Mastercard,
23 what is Mastercard? I think most of you are familiar
24 with it as a brand name, but you do not necessarily
25 know what we do.

1 So we are a network, we are a technology
2 provider. We connect your banks, your consumer banks
3 to the merchants' banks and, therefore, enable
4 cardholders, you who are holding an account, to
5 actually make a purchase with a merchant. But you are
6 not our customers, the merchants are not our customers
7 for the most part for our core network. Instead, you
8 are indirect customers.
9 So as part of that, we are the pipes that
10 connect everybody to each other. So we do not issue
11 the cards. That is one of the biggest fallacies
12 people have about Mastercard. Instead, see the logo
13 on the front? That says Citi. And if most of you
14 pull out your cards, you will see it has the bank's
15 name who you have the relationship, who you give your
16 personal data to. Instead, you have these things on
17 the back that says "bug," which we call acceptance
18 marks, that says if you go into a store, this will be
19 accepted.
20 So what does that mean from a data
21 perspective? From a data perspective, I do not have a
22 data relationship with consumers. Instead, what I
23 have is I get enough data to process a payment. What
24 is that? That is an account number, the amount -- the
25 time of the transaction, the total amount of the

1 transaction and the merchant.
2 So actually, I would like to say 40 years
3 ago, somebody had the foresight to actually do some
4 privacy by design because I do not have your name.
5 I do not need your name to process a transaction.
6 I do not know what you actually buy. I do not need
7 that to process the transaction. Instead, the bank
8 gets the information. The bank says, oh, $50, Yael
9 has $50 in her account, yes, she does, and she is
10 waving her hands and so, therefore, I will approve
11 the transaction. Well, I think Leisl is actually
12 Leisl and I will approve it because I actually think
13 it is her actually making the transaction. Or if I
14 think it is some fraudster, I will not approve the
15 transaction.
16 So what do we do? So as a result of that,
17 we see 55 billion transactions or so. The number
18 keeps growing exponentially, thank goodness, for our
19 jobs, of transactional data. So what do we do with
20 that data? Well, I will tell you one of the great
21 things that we do with it is we innovate. We are
22 constantly developing new products and solutions, and
23 one of the most important products and solutions that
24 we develop to help all of us, me inclusive as a
25 cardholder, is to protect all of us from fraud.

1 So what does that mean? So, historically,
2 where all of the data's been coming from that amount
3 of transaction, time of transaction, that is happening
4 when you are at the cashier in the old days, right,
5 and you would swipe your card. It would come through
6 and we would see it and then the bank would have to
7 authorize that transaction. So you would stand there
8 and hopefully wait only the five milliseconds where
9 you are saying, it is approving, approving.
10 So during that time, we have tools that
11 enable us to do determinations and to start doing risk
12 scores to say, do we think this is fraud? Do we not
13 think this is fraud? Now, those have evolved over
14 time. In many cases, they used to just be rule-based,
15 simple if/than. Nowadays, we use AI to do it.
16 So as we have grown our models, as we have
17 grown our technology, we are able to protect people
18 more and more. And another great thing about this, as
19 the rest of the world has moved to adopting cards and
20 payments through accounts like that, then we have
21 enabled protections against fraud for those consumers
22 across the world.
23 Over time, though, we say, okay, this is our
24 basic data set. How else do we help improve the fight
25 against fraud? Because it is an arms race. There are

1 constantly new players coming in trying to steal data.
2 There are constantly new players trying to come in and
3 make runs against banks. So we work with the banks to
4 say, hey, how do we help you here?
5 So in many case, we have worked with the
6 banks to segment you. So we use the data to help
7 determine, hey, here are classes of consumers and this
8 is how they behave, and based upon those
9 classifications, this is what we think fraudsters look
10 like. This is what we think your people look like.
11 So does this help make determinations? Does this help
12 you reduce fraud?
13 Another service we work with them on is to
14 actually get some information from them or have them
15 get information. So in e-commerce situations, a
16 merchant can pass them the name and they can actually
17 also check that name. Now, Mastercard does not have
18 to get the name. Instead, we are enabling the pipes
19 that allow for passing the data.
20 Then as technology has evolved, we have
21 evolved to new payment forms. Now, who has Apple Pay,
22 Android Pay, and who has used it? Very nice, simple,
23 easy way. And I am sorry, Garmin, you guys can use
24 it, too, on Garmin. So we helped build the backbone
25 for that, so you can thank the payments industry for

1 enabling you to just put it on your phone.
2 But when you put it on your phone, what are
3 you doing? Most of you, I think, have gone through an
4 authentication experience. You are providing your
5 name and address so that Apple, in one case, just
6 sends it through to your bank and your bank then
7 confirms that is you. So you are authenticating
8 yourself to your device.
9 Mastercard just needs that data for a very
10 short time period. We do not really need to retain
11 it. I do not need to continue to authenticate you.
12 So, again, privacy by design, it happens once. But,
13 now, you get to be authenticated to your phone. And
14 so, now, I have an additional way to say hey, this
15 phone actually is Andrew and, therefore, I get that
16 little flag that says, hey, this is great, Andrew just
17 got authenticated to his phone. Mastercard does not
18 know Andrew; Mastercard might know that it is Apple,
19 Apple device or a Garmin device in the case of Garmin
20 that the payment occurred. So that is how
21 authentication might work.
22 The other way with e-commerce merchants is
23 we work with e-commerce merchants and mobile
24 merchants, m-commerce merchants, to say, hey, guys, if
25 you give us more data or enable some collection of

1 data, we can help you fight fraud even more
2 effectively. So imagine if I'm only seeing the same
3 account number against the same iPhone. Gee, that
4 persistency tells me that there is a reduced risk of
5 fraud here or if I see the IP address as from certain
6 parts of Eastern Europe, that are known for high
7 fraud, I can also say high, high risk of fraud here.
8 So those are the kinds of things that we are
9 doing to try and help fraud. This is how we -- to
10 fight fraud, not actually help advance it, sorry. And
11 so we are constantly looking at new ways to use data,
12 to look at new data sets, to build on data sets, but
13 as we are dong that, we are trying to minimize the
14 data sets we have. If you do not have the data, you
15 cannot lose it. If you do not have the date, you
16 cannot accidentally abuse it. So, therefore, a lot of
17 privacy by design and data minimization as we are
18 doing product development, but all in furtherance of a
19 good cause to help to protect all of you from fraud.
20 (Applause.)
21 MR. COOPER: All right, thank you, Andrew.
22 And, you know, we have about 23 minutes left of
23 discussion here. I heard a nice panoply of the uses
24 of data across different industries.
25 One of the things I heard a couple of

1 panelists mention and I would like to get others who
2 have not weighed in on this to speak about, is when we
3 think about big data, you know, one of the things that
4 sometimes sets apart big data from just normal data is
5 that you are looking -- the analysis that is performed
6 on it often is looking more for patterns that emerge
7 that you could not see with smaller data sets.
8 You are looking for associations, as opposed
9 to, when you think about it, sort of normal in
10 economics -- you know, Florian mentioned this in his
11 presentation earlier this morning about the, you know,
12 kind of gold standard of causation -- and you are
13 looking in control groups and figuring out.
14 So one of the questions I had -- and since
15 we have not heard from Florian in a while, I will
16 start with him, but anyone else can jump in -- is, you
17 know, in general, what is the relative importance of
18 both looking and finding causal versus associations,
19 and sort of related to that, when you think about big
20 data, what is more valuable? And I think I already
21 know the answer you are going to give as seen in your
22 presentation. But what is more valuable having a good
23 team or knowing how to ask the right questions or
24 actually having access to a large and comprehensive
25 data set, actually having access to big data? So what

1 is more important in that?
2 And I will start with Florian, but I would
3 like anyone else to jump in.
4 MR. ZETTELMEYER: I think on your first
5 question about kind of what kind of data is the most
6 useful, I would simply say that it is incredibly
7 context-dependent. Roughly speaking, I think of
8 analytics creating three things. It can enable
9 business initiatives. Like if you think about
10 personalization, that is really an enablement
11 function. You are creating something that allows you
12 to achieve an outcome. A lot of the things, for
13 example, that Morgan was talking about I think fit in
14 that area as well, as well as a lot of the things that
15 Liz was talking about, design-abling things.
16 Then I think the second big use is that it
17 enables you to basically come up with ideas. That is
18 what you were talking about, about large data sets
19 where you can look at correlational patterns and see
20 whether you can come up with ideas from that.
21 And the third one for me is that data allows
22 you to evaluate whether things that you are doing are
23 reasonable or not and whether they work or not. So,
24 for example, my first talk this morning was really
25 about evaluation. It was like, you know, is this ad

1 working or not? It did not help you come up with the
2 ad, it did not help you necessarily kind of enable the
3 ad. That is what these -- obviously, these targeting
4 mechanisms do.
5 So I think it just depends completely on
6 what the purposes are. I think one of the mistakes
7 sometimes people do is to think too narrowly about
8 what uses of data exist. And they are very different
9 from each other and you need very different data.
10 Sometimes it has to be causal. In many cases,
11 causality is not at all interesting or required. It
12 is just a matter of what you are looking for.
13 On the second question of what both you
14 need, I actually think that data and skill teams are
15 complements and not substitutes. So to the degree
16 that you have better data, having the ability of
17 asking great questions suddenly becomes more valuable
18 to a particular firm.
19 MR. COOPER: Okay, thanks. Morgan, down
20 there?
21 MR. REED: So it was interesting. You know,
22 I think that it is one of those that are intertwined.
23 But I know that there are some folks in the audience
24 here who are more specialists in this, but I think
25 some of the things that have been revealed through

1 some of the criminal justice reform analysis of big
2 data have been profound and a bit disheartening, but
3 they go to this value of -- what is the old phrase?
4 That quantity is a quality all its own. And sometimes
5 in data the ability to see large shifts or check for
6 some various effectiveness, as Florian talked about,
7 is almost impossible because to separate the signal
8 from the noise ratio is too hard.
9 And so I think when you say, well, what is
10 the most valuable aspect? Skilled teams, data set,
11 size, those elements of it, I think they are fairly
12 intertwined, but I would recommend that everybody take
13 a look at criminal justice reform questions where big
14 data has been used to show some, like I said, fairly
15 depressing things about if you want to go before a
16 judge, make sure you do it at this time and not after
17 -- you know, before lunch but not when they are
18 hungry. The fact that hunger seems to have more of an
19 impact on whether or not you go to jail as opposed to
20 what you have actually done as a crime.
21 I do not think you can reveal that without
22 big data sets. And then as you point out, you can
23 reveal it with big data sets, but you have to be able
24 to ask the right question.
25 MR. COOPER: Mark?

1 MR. MACCARTHY: So I thought the magic word
2 in the last comment was context-dependent. So do you
3 need large data sets or small data sets? You know, it
4 is like it depends on what you are using it for.
5 Sometimes you need a large data set to get the result.
6 As I think you mentioned earlier, there are studies
7 that show that these effects of size diminish after a
8 certain point and you can add more data to the data
9 set and you do not get anything new. So there are
10 diminishing returns.
11 And also in a context-dependent sense,
12 whether the information you have is valuable for a
13 long period of time or whether its value decays
14 quickly depends on the context you are operating in.
15 If you have search information, that decays very, very
16 rapidly. You know, someone may be identified as being
17 interested in a vacation in Maine in August, but you
18 better not send him an advertisement for that in
19 December, he probably is not interested.
20 But on the other hand, medical information
21 might be very valuable years after the data has been
22 collected. The analysis can still be done even though
23 the information is not fresh and insights can be
24 gathered even though data is not last year's data. So
25 I think it does depend on the context and we have to

1 be very, very careful not to make broad
2 generalizations about how valuable is data over time
3 or whether large data sets are better than small data
4 sets. You have to look at the context in which the
5 information is being used.
6 MR. REED: I want to amend my answer with
7 one thing that Mark brought up that is really
8 important. Mark said something really important. He
9 said, “but medical data.” And here is the thing you
10 heard in what Christopher said and what Liz talked
11 about and what Mark kind of brought up, which is we
12 are not 100 percent sure what is medical data. When
13 we are trying to figure out whether or not there is a
14 cancer cluster, I may need to look at other factors
15 that might not be obvious, that might not have fit
16 into our current understanding of what is medical data
17 in terms of how the FDA judges our product.
18 So I think, Mark, you were spot on and I
19 think it ties into with what you heard from
20 Christopher and Liz and others. We are not exactingly
21 sure of all of it, but we want treatments that reflect
22 us as a holistic person not merely the data that is
23 contact in our EHR. So I think it is a good point,
24 Mark.
25 MR. COOPER: Florian, you had a quick

1 followup?
2 MR. ZETTELMEYER: Yeah, I just wanted to say
3 one more thing about the complementarity of the data
4 and the team. A lot of firms for internal processors
5 are using data to basically improve decision-making.
6 So one of the interesting things about this is that
7 the better decisions get as a result of having used
8 data, the less variation exists in business processes
9 because the data was used in order to optimize those
10 decision processes. This is why, you know, we use the
11 data in the first place.
12 What that also means is the data is getting
13 less useful over time because now you have less data
14 variation, and as a result of that, the importance of
15 the team is to know when to inject more variation into
16 the data in order to be able to still measure what is
17 going on. In other words, you say do you have
18 experimental design and variation of data and thinking
19 of manipulating or rather designing or varying data as
20 a strategic imperative is incredibly important. That
21 does, at the moment, at least require some teams to
22 set that up.
23 MR. COOPER: Anyone else like to jump in?
24 MR. REISKIND: I think I will just
25 reenforce. I have had very personal experiences

1 dealing with geospatial data lately, because a lot of
2 our analytics are based upon where a merchant might be
3 located, as well as your cell phone might be located.
4 And some of the analytics can be worked off of very
5 crude locations, like especially outside the United
6 States, quality of data is kind of limited. There are
7 not postal codes, there is only one city in the entire
8 country, things like that. And you have to work with
9 that as a data quality issue that you cannot overcome,
10 and so it limits some of the things you can do.
11 But there are things you can do with that
12 data, but they may not be as good as you want to do.
13 So, for example, to tie my cell phone to my physical
14 location, my cell phone to my physical location where
15 I am making a spending purchase would be our nirvana.
16 And in some cases, we can get to that nirvana to prove
17 my iPhone is where I am making an expenditure is a
18 great thing, because them it proves I am not a
19 fraudster. But in many cases, you cannot get there.
20 So you have to mediate what your innovation
21 is and what you are trying to do based upon the
22 quality of the data that you are dealing with as well
23 as the skill of the data scientist and the tools you
24 have to work with the data. Geospatial data is a very
25 unique data set -- sorry, postal addresses tend to be

1 not very useful for analytical purposes. You need to
2 take 4100 Yuma and actually turn it into a lat-long
3 for analytical purposes to stick it in a model. 4100
4 Yuma will not work very well in a model, as can you
5 imagine in a mathematical algorithm.
6 So, therefore, geospatial data sets at least
7 need that level of transformation and, yet, that is
8 only as good as the maps are in Third World countries
9 or underdeveloped countries in many cases. So that is
10 just an example. Like it depends on the data set, it
11 depends on the tool, it depends on the use case. It
12 is all very context-driven.
13 MR. COOPER: I want to switch gears a bit
14 here. Liz talked about this in her remarks, about
15 regulation that we see, the GDPR and the recent
16 California privacy law, that both -- what I would be
17 interested in hearing from all of you is to what
18 extent do you see either of those types of regulation
19 impacting your use of data and how might that
20 ultimately impact consumers. So anyone who wants to
21 jump in.
22 Mark, you had your hand up first.
23 MR. MACCARTHY: So I think it depends a lot
24 on whether you are dealing with a large company or a
25 small company. The compliance burden for both

1 California and for GDPR, for large companies, if it is
2 the kind of thing that they can do, and with
3 sufficient resources, they can find a way to comply,
4 they will be able to do it. I think one of the
5 previous speakers talked about 800 hours of compliance
6 work that was put into getting into compliance.
7 For larger companies, like many of the
8 companies in my trade association, that is doable.
9 But for many of the smaller companies -- and we have
10 700 companies in my trade association -- many of whom
11 are very, very small and they would love to operate
12 globally. For them, the choice came down to enormous
13 compliance costs for operating in Europe versus not
14 operating in that market at all, and for them, it was
15 an easy choice.
16 So I do think we have to pay very, very
17 close attention to the compliance costs that are
18 imposed on businesses. If something is really needed
19 to protect consumers against real harm, then you got
20 to do it and people pay the compliance costs. But if
21 it is just a lot of extra processes, you know, put in
22 there to validate that you are doing the right thing,
23 then there may be less benefits from those compliance
24 costs than we would like.
25 MS. LOPEZ-GALDOS: I completely agree with

1 you, Mark, and I would like to add just a tiny bit
2 there, which is that resources that are taken to
3 comply with the laws because, obviously, if we adopt
4 regulations, companies are going to comply with them,
5 those are the resources that the smaller companies are
6 going to stop investing in innovation. So we also
7 have to look into the actual effects of the need to
8 comply with the law.
9 And it is certainly the case that big
10 companies can comply with those new laws much easier
11 than smaller ones. So I think that is a very
12 important point that Mark was making.
13 MR. REED: And it was worth noting that Liz
14 mentioned not 800 hours, she said, hundreds of
15 person-months. So I want to remind everybody that
16 here is the part that is so cool. Earlier, I talked
17 about two billion people having access. My smallest
18 companies are global players. Our current board
19 president has an app -- kind of a cool app, he has 2.8
20 million users in about 117 different countries. He is
21 a one-man shop in Oregon.
22 My example I always use is my literal
23 smallest company, Ann Adair's company that makes kids
24 apps, she is a music teacher that is a part-time coder
25 with her kid and her husband and has a whole slew of

1 really cool kids apps. She is a global player with
2 hundreds and hundreds of thousands of users. So Liz's
3 point about hundreds of person-months to comply with
4 GDPR has a real implication.
5 And I will dig down to one area of
6 specific that gets into the business of big data.
7 If you are not familiar, right prior to the launch of
8 the GDPR, the Article 29 working party released a
9 letter directed at ICANN specifically about the
10 ability to using the word "including" in your terms of
11 service. And this is always an awkward thing to bring
12 up because everybody is essentially ignoring this
13 letter.
14 In this letter, ICAN was told, you may not
15 use the word "including" because to use the word
16 "including" means you are not being complete,
17 comprehensive, and explicit. And here is the problem.
18 We are on a panel of the business of big data. How
19 can I cover all the algorithmic learning that I am
20 going to do and be explicit and comprehensive when I
21 quite literally do not know the answer of where the
22 data might take me and back to causal and correlative
23 effects.
24 So I think there are moments where well-
25 meaning regulators will put language in like that and

1 then the outcome, from a data science perspective, is,
2 well, I do not know what the outcome will be, so how
3 can I be comprehensive and explicit? So I think we
4 need to be cautious about just jumping on board and
5 say that the U.S. version of GDPR needs to plug and
6 play. I think we need to ask real questions about how
7 it will impact good use of big data to solve real
8 problems that people have. So hundreds of
9 person-months plus loose regulatory language will have
10 an impact.
11 MR. COOPER: Did you want to jump in?
12 MS. HEIER: Yes. So just to kind of
13 clarify, right? I said 800 person-months of effort
14 and that is really not correlated to the number of
15 users we have or the number of countries we operate
16 in. We have 30 years' worth of devices, services and
17 data that we had to bring up to compliance. So it
18 does not really matter necessarily size of the
19 company. It is really your offerings, right?
20 So as you said, it could be one person that
21 is operating out of their garage part-time, but
22 operates and has lots of data. Their cost of
23 compliance is going to be much different than ours.
24 MR. COOPER: That is a good point. Anyone
25 else like to jump in before I get into some questions

1 from the audience?
2 (No response.)
3 MR. COOPER: Okay. So this one is directed
4 at Mark, so I will let you take first stab at it, but
5 open it up for everyone else. And it has to do with
6 you talked about credit scoring, how using alternative
7 data and big data methods can actually lead people who
8 do not have credit lines to have lines and be scored
9 or are unscored and be scored.
10 And this question says, perhaps that makes
11 sense in a credit-scoring situation, but sometimes if
12 you are training data set -- if you are training these
13 algorithms with historical data, in other contexts,
14 perhaps, they can ingrain bias. So is that something
15 that you should worry about in the context of big data
16 and AI?
17 MR. MACCARTHY: Yes. Actually, credit
18 scoring is one of the areas where they have had
19 experience with bias and statistical discrimination
20 going back for generations. The credit scoring world
21 is under a legal obligation to avoid the
22 discrimination in lending. The fair lending laws
23 require all of the credit scores that are used in that
24 area to pass a disparate impact test, which means they
25 have to look carefully at whether their algorithms

1 have an adverse effect, a disproportionate adverse
2 effect on minority groups. And if they do, they have
3 to ask themselves, what is the particular purpose they
4 are involved in that makes this disparate impact so
5 important? And if they have a legitimate business
6 need, then they have to also ask themselves is there
7 another model, another credit-scoring model that will
8 achieve the risk reduction that they are looking for
9 with less of a disparate impact?
10 So all of the credit scoring models have to
11 pass that test if you are in the business of producing
12 one of those models to people who buy it or people who
13 will be examined by federal regulators for compliance
14 with the fair lending laws. Now, if you happen to use
15 machine learning, you know, in that context, that is
16 not a get-out-of-jail-free card for getting rid of
17 discrimination charges. It does not work to just say,
18 well, I used artificial intelligence so I do not have
19 to comply with the fair lending laws anymore. So the
20 new techniques are as much covered under the old laws
21 as the old techniques were, and in that particular
22 case, there really is a regulatory requirement to
23 avoid discrimination.
24 MR. COOPER: Would anyone else like to weigh
25 in on that in general? I mean, I think related to

1 that, a bigger-picture question is, in general, we
2 think about using big data, using analytic methods or
3 the predictions. Are they more -- we have to look at
4 what the alternative is. Are they more or less
5 discriminatory than what the alternative would be or
6 more or less accurate than what the alternative would
7 be? And I just wonder if -- this is kind of related
8 to the question from the audience that was thrown out
9 to Mark. Does anyone have any thoughts on that?
10 MS. LOPEZ-GALDOS: I can jump in. I think
11 one of the keys is going to be able to explain, and AI
12 models are going to have to be able to explain how
13 they operate. So definitely the laws are there, the
14 principles that need to be protected are there. The
15 fact that you use an AI or machine-learning
16 methodology is not going to change your obligations,
17 as Mark said, and the difference is that we are going
18 to have to determine what the explainability of those
19 AI models are going to be to be able to prove that we
20 comply with the laws. So I think that is going to be
21 key.
22 MR. COOPER: All right. We are rapidly
23 running out of time, but here is another question from
24 the audience that says, if we look at data as an
25 asset, how should companies treat this from an

1 ownership perspective? Should it be treated like
2 intellectual property? Should consumers have any sort
3 of ownership interest in this? So how should we think
4 about big data in this context?
5 MR. REED: Well, there are multiple stages.
6 We have rules governing your health data. Your health
7 data is your data. But the question is once it is
8 manipulated, once the physician has put additional
9 work and information into it, then where does it
10 stand?
11 The work product of the physician is
12 valuable and valued. So how do we work with that
13 becomes a real question. When it comes to something
14 most people do not know -- we have not talked about
15 HIPAA at all, but the P in HIPAA stands for
16 portability not privacy. So a lot of the questions
17 about big data are very interesting because your
18 health data in particular is something that there is a
19 push to make it portable so you can move it from place
20 to place so the physician is well armed in order to
21 treat your disease.
22 The question was interesting and you touched
23 on it earlier when Marianela was talking on the
24 intellectual property question. The explainability
25 and transparency of the algorithm also gets very

1 interesting in so much that what you have trained and
2 what you have learned is also a work product of your
3 company and might be protected. So how do you
4 separate the data sets from the work product? If the
5 data set -- if the work product is actually trained
6 off of those data sets, then which thing is the asset?
7 I think the reality is healthcare is, in a
8 weird way, almost easier because there has been this
9 kind of agreement across the industry that your health
10 and your specific healthy information is yours, the
11 patient's property. But it does get interesting into
12 the question of what is the value of the work product
13 that is created off of that data set and where does
14 that set in the realm of intellectual property.
15 MR. COOPER: Mark?
16 MR. MACCARTHY: Yes, I think the ownership
17 lens is the wrong one to bring to bear in this kind of
18 circumstance. I mean, most information is about more
19 than one person. I mean, if I bought something from
20 you, then you sold something to me, and so the
21 question of who owns the data is an attempt to import
22 sort of property law into that circumstance and it
23 just does not help you very much in trying to figure
24 out what the right thing to do is.
25 If I own the data, does that mean I can

1 destroy any copy of it anywhere, any business record
2 in the world I can sort of destroy because it is mine?
3 Well, that does not make any sense. So I think you
4 might as well go directly to the data protection rules
5 and regulations and the responsibilities on both
6 parties to try to figure out what the right thing to
7 do is, rather than say, I am going to define who owns
8 it and that will end the problem because now I know
9 who owns it.
10 I think you will not be able to solve the
11 problem of determining the right owner, so I think you
12 just have to go to what are the rules, what kind of
13 consent needs to be given, what kind of access is
14 there, what kind of portability rights are there, and
15 those things really take a lot of careful and hard
16 thought, and you cannot really solve those problems by
17 saying, I fixed the problem, I decided who owns the
18 data.
19 MR. COOPER: Okay. Florian and then Liz.
20 MR. ZETTELMEYER: So, Mark, I agree with you
21 that that is true on the regulatory side, but, I mean,
22 in terms of data usage on the company's side, that is
23 a problem that shows up all the time, and particularly
24 in disintermediated industries like, you know, do you
25 own your data or does the physician own the data or

1 does the retail own the data or does Procter & Gamble
2 own the data and what are you allowed do with it, et
3 cetera. So, I mean, it is an issue that companies
4 have to grapple with. It may not be useful from a
5 regulatory point of view, but it is certainly
6 something that is pretty omnipresent in this data
7 world.
8 MR. MACCARTHY: I think you have picked on
9 the key, which is what are companies allowed to do
10 with it. That is the question. You do not resolve
11 that by saying, I know who owns it, therefore, I know
12 what use requirements there are. I think you have to
13 go directly to the use restrictions and constraints
14 and who has what right to do what with it.
15 MR. COOPER: Liz? This will be the last
16 word.
17 MS. HEIER: Well, just to reiterate what I
18 said in my statement, Garmin believes that the data
19 belongs to the user and the customer. They give it to
20 us to help enhance their experience, to give them new
21 data points they would not have on their own. So we
22 have really formulated, you know, our data privacy
23 program around that user-centric focus.
24 MR. COOPER: That is perfect, zero, the
25 clock is at 30. So well-timed.

1 THE IMPACT OF GDPR ON EU TECHNOLOGY VENTURE
2 INVESTMENT
3 MR. STIVERS: Okay. I think we are going to
4 go ahead and start the afternoon session so we can
5 keep our somewhat amazing track record of staying on
6 time for this hearing. So thank you to OPP and the
7 FTC staff for having kept us on track.
8 I am Andrew Stivers. I am the Deputy
9 Director for Consumer Protection in the Bureau of
10 Economics, which just means that I am basically in
11 charge of the Consumer Protection economics mission at
12 the FTC. I am delighted to basically just introduce a
13 series of really good speakers this afternoon. So I
14 am going to step out of the way and we are going to
15 start with Liad Wagman, who is a Professor at the
16 Illinois Institute of Technology in the Stuart School
17 of Business.
18 Liad?
19 MR. WAGMAN: Thank you again for having me
20 here today. This is joint work that is fresh off the
21 copy machine pretty much with Ginger Jin and Jian Jia.
22 We have been in a mad dash to complete it over the
23 last several weeks.
24 Basically, we looked at GDPR and we asked
25 ourselves where would we notice an impact right away.

1 And the answer we came back with is that investors are
2 likely to internalize the effects. So we thought the
3 law was passed a couple years ago, back in 2016, maybe
4 we should notice an effect then because investors
5 would form expectations. The thing is, not much was
6 seen and we were wondering why.
7 Looking through the news events, we saw that
8 as recently as early this year, more than half of
9 mobile applications are not GDPR-ready, and
10 announcements very close to the implementation date,
11 to the enforceability date of May 25th, kept pouring
12 in. The top firms, the top platforms started
13 releasing their rules.
14 Apple removes apps to share location data
15 without consent, updates their privacy terms.
16 Facebook says that businesses may want to implement
17 code that creates a banner and requires affirmative
18 consent. Each company is responsible for ensuring
19 their own compliance. You are all on your own.
20 Shopify updates its app permissions for
21 merchants/developers. They need to implement them.
22 Google releases consent SDK for developers, these
23 software development kits, just a day before, the
24 eleventh hour before the enforceability date, and then
25 GDPR takes effect. So we kind of understood this all

1 came to this implementation stage of the regulation
2 and so the effect should be noticeable after that or
3 as this was happening.
4 So this is sort of our motivation, you know,
5 GDPR has a massive overhaul of data regulation in the
6 European Union and anyone who services the European
7 Union. That includes data management; auditing and
8 classification; data risk identification; risk
9 mitigation; interfaces for users to obtain their own
10 data to provide opt-in consent and to request deletion
11 of their personal data. Firms are required to train
12 or hire qualified staff or they face severe penalties
13 that are up to 4 percent of their annual global
14 revenue.
15 Bloomberg, shortly after, said, 500 biggest
16 corporations are on track to spend a total of $7.8
17 billion to comply. Now, based on earlier work, we
18 already knew that compliance costs are not incurred
19 equally by firms. Smaller firms tend to take a bigger
20 burden, at least in relative terms. And the other
21 effects we know from theoretical work is that
22 compliance cost will shift some of the innovation
23 activity from smaller firms into the bigger firms.
24 And the reason, especially for tech, that
25 this happens is because larger firms already have the

1 infrastructure in place for R&D. They have the
2 infrastructure in place for internal innovation. So
3 when entrepreneurs decide to pursue an idea, they have
4 the option of pursuing it internally or pursuing it as
5 a startup externally, as a venture. When they face
6 that choice, they look at the cost, and when the cost
7 of pursuing it on your own increases, your incentive
8 to stay inside and either innovate or not increases.
9 And so the overall, at least, theoretical effect is
10 that innovation is reduced and more innovation happens
11 inside bigger firms.
12 So the bottom line for us was who is better
13 to assess what really happens than the actual
14 investors who are putting their money where their
15 mouths are, that are actually investing in those
16 firms. So once these policies were rolled out, we
17 figured compliance costs are going to be realized,
18 especially for the smaller ventures because they rely
19 on the larger platforms' policies for compliance, for
20 who bears the liability for violation, and so forth.
21 So that is the general idea.
22 Now, we wanted to get comprehensive venture
23 data. It is impossible to get it all in one place,
24 but one of the main databases for venture data that is
25 not a complete universe, but it is pretty good, is

1 Crunchbase. So we collected venture data from
2 Crunchbase from last summer, July 2017, until the end
3 of September, this year. So it is really, really
4 recent.
5 This data comprises firm information, the
6 firm location, the category it operates in, its
7 founding date, the dates on which it raised money and
8 a range, a lower bound and upper bound, on the number
9 of employees it has. Think 1 to 10, 11 to 50, 51 to
10 100, something like that.
11 Now, it also comprises information about
12 each individual financing deal. That includes the
13 size and the date of the deal, which stage, was it a
14 seed deal, a Series A, and so forth, which investors
15 participated, and the dollar amount obviously of the
16 deal.
17 So just to give you an idea of what the data
18 looks like and to convince you that it is good data, I
19 created some pictures to kind of summarize it. So
20 these first four pictures show the average number of
21 deals per week in the U.S. and in the EU. You can see
22 the U.S. has a larger number by a factor of two or so.
23 The median dollar amount in millions raised per deal
24 is about a million and a half for the EU and three
25 million for the U.S. You can see the average firm age

1 is more or less similar and the average number of
2 investors that participate in a deal is somewhat
3 higher in the U.S.
4 If we look at the composition of firm ages
5 in our sample, you will notice that about half of them
6 are the very youngest, the zero to three years old
7 ventures. And the rest are distributed more or less
8 similarly between the U.S. and the EU.
9 So if we dig deeper into these age groups,
10 you can see that the average amount raised per deal is
11 growing the older the firm is. So the youngest group
12 raised the least, they mostly participate in seed
13 rounds and Series A, Series B rounds, and then grows
14 from there. These are averages; they are not medians,
15 so the amounts are a little higher.
16 Now, if we look at the total number of
17 deals, most of the deals happen for those young firms.
18 They have smaller deals, but they have a lot more of
19 them. And we are talking thousands of deals in just
20 one year of data, a little over a year. And if we
21 look at the median amounts raised per deal, you notice
22 that, again, they grow in the firm's age, and this is
23 kind of indicative that the distribution of those
24 amounts is skewed. The median is smaller than the
25 average.

1 So if we want to dig deeper into the types
2 of deals that are happening, I hope I have convinced
3 you by now that this data is pretty granular, but it
4 goes further than that. You will see that for those
5 youngest firms, those zero to three year old firms,
6 most of the deals happen on this large circle which is
7 the seed round. Those are the smallest basically
8 rounds that mainly comprise angel investors and
9 amounts of a few hundred thousand dollars a deal.
10 Then it goes from there. It goes to Series
11 A, Series -- bridge rounds A-B, and others. So on the
12 horizontal axis here, you have the firm age; on the
13 vertical axis, you have the average dollar amount for
14 deals of that type. And then the larger the circle,
15 the more deals we see.
16 As we move to older firms, you will notice
17 that the bubbles start floating up as the deal amounts
18 increase and there are fewer deals so the bubbles get
19 smaller. We can go to the older group and they keep
20 floating up, the age obviously increases, the bubbles
21 get smaller. And we could go to the oldest group and
22 they keep floating up.
23 So in it terms of where those deals are
24 happening, this is a heat map of U.S. states and the
25 EU member states. We include Britain in the EU

1 because it was still part of the EU as of the time of
2 GDPR's rollout. The EU firms are affected by GDPR
3 just as much. In fact, the U.K. adopted its own
4 GDPR-like law.
5 You will see most of the deals happen in
6 California, happen in the U.K. In terms of the dollar
7 amounts that go in, it is a pretty similar situation.
8 Most of the dollar amounts go to the U.K. and
9 California and Germany picks up some investment
10 dollars as well.
11 So our observation level here is divided
12 into a state, where a state is either a member state
13 in the EU or a state in the U.S. So we look at least
14 at the aggregate level at states.
15 In terms of time, we look at weeks.
16 Investment per week, per state, per technology
17 category. I will talk about categories in a second.
18 At the deal level, we look at individual
19 deals. So I hope this was convincing at least in
20 terms of the granularity of the data we have.
21 Let me give you some idea of the trends
22 here. This is for the number of deals per week
23 comparing the EU and the U.S. The U.S. is the red
24 line; the EU is the blue line. You notice that they
25 track each other pretty closely. It seems to be a

1 common trend and GDPR takes effect in late May this
2 year, and there seems to be some change going on.
3 Now, you might argue, oh, this is the European summer
4 vacation happening right after GDPR takes effect, but
5 we do not see a similar thing in the summer of 2017.
6 We dig deeper into the deal per week per
7 state per technology category level. You will notice
8 that this gap becomes easier to spot, this gap that
9 happens between the red line representing the U.S.
10 trend and the blue line representing the EU trend.
11 And there is a drop that happens after GDPR takes
12 hold.
13 We could look at variations of this of the
14 dollar, for example, raised per week, and see the same
15 thing. We could go further and look at the dollar
16 raised per week per state per technology category, and
17 again, we can see the same thing. And we could look
18 at the dollar amount raised per deal and, again, we
19 see something similar taking shape.
20 So our next objective here is to quantify
21 this effect, to look empirically at what is going on.
22 Our methodology is what is called difference-in-
23 difference. So what we do is we find the difference
24 in the U.S. from the pre-period, before May 2018, and
25 the post-period, after May 25th, 2018, and we do the

1 same thing for the EU, and then we take the difference
2 of the differences.
3 So we have a couple specifications. At
4 least at the aggregate level, we use Tobit for the
5 total dollar amount raised per week per state and we
6 use Poissno for the number of deals per week per
7 state. We use macroeconomic controls, like
8 unemployment, consumer price index, GDP. We even
9 included exchange rate. That did not change anything.
10 And a specification is what you would
11 expect. We are just looking for the effect of the
12 rollout of GDPR. We use time and state, country fixed
13 effects for the EU, and at the deal level, we use a
14 log linear specification because of these outliers
15 that we have where we see the average is much larger
16 than the median and this helps control for that.
17 At the deal level, we also include the
18 deal-specific controls like the age of the firm, the
19 funding stage of the deal, technology category, things
20 like that. And in terms of technology category, we
21 break it down into two categories. One is healthcare
22 and finance, and the other is everything else. The
23 reason we focus on healthcare and finance is because
24 the U.S. has existing laws in those sectors;
25 specifically, the Gramm-Leach-Bliley Act, GLB, for

1 finance and HIPAA for healthcare. So we would expect
2 maybe to see something different about that category,
3 that grouping of healthcare and finance.
4 The other reason we divide it into these
5 categories is because it creates a valid sample in the
6 sense that every state has some activity in those
7 categories.
8 So in terms of results, we see an effect on
9 the dollar amount raised per week per member state per
10 category that is substantial. Across all EU ventures,
11 that dollar effect is $3.38 million per week per state
12 per category. For zero to three-year-old ventures,
13 the effect is almost a million dollars.
14 Now, in terms of the number of deals, we see
15 a significant drop, a drop of about 17 percent for the
16 number of deals per week per category per state. The
17 figure represents the average amount, just to make it
18 easier to kind of relate to. And we see a similar
19 drop for those youngest ventures, those zero to three-
20 year-old ventures. What this means is that those
21 firms have less of a chance to secure a successful
22 deal which could mean that fewer of them come to
23 fruition.
24 In terms of the dollar amount per deal, that
25 also drops. Those drops are pretty significant in the

1 overall sense because some of the later deals are very
2 sporadic. When we zoom in on the zero to three-year-
3 old ventures, the drop there is 27 percent.
4 Overall, we see two effects. We see an
5 effect at the extensive margin in terms of fewer deals
6 taking shape after GDPR takes hold, and at the
7 intensive margin, in terms of fewer dollars invested
8 per average deal.
9 Let's talk about some of these categories
10 more specifically. So in terms of healthcare and
11 finance, we see a similar drop in the number of deals
12 of 18.8 percent. We see a drop in the aggregate
13 amount raised per week per state of $5 million. The
14 average amount invested per week is $30 million. And
15 we see a huge drop in the amount invested per deal, on
16 average.
17 Now, interestingly enough, we see similar
18 changes for all other categories. We do not get a
19 significant effect on the aggregate dollar amount
20 invested per week because that pool of categories is
21 just too widely spread. It is too broad. So we are
22 not able to identify that effect, but otherwise it is
23 somewhat similar. This is surprising because you
24 would think that healthcare and finance would be
25 different since the U.S. has existing laws.

1 Now, what we get out of it is that maybe
2 GDPR is really transformative in the overall sense
3 across categories. It doesn't matter if there are
4 existing laws; those laws are old. They are outdated.
5 There are systems in place already to handle those
6 laws. Whereas GDPR is new, is fresh, needs new
7 systems, new compliance costs.
8 Now, zooming back into those zero to three-
9 year-old ventures, those nascent ventures, those
10 startups, the effect there is pronounced. There are
11 19 percent fewer deals happening. There is a decrease
12 in the aggregate dollar amount invested per week and
13 there is a drop in the dollar amount invested per deal
14 on average. That is, to me, concerning. And at the
15 same time, we do not know if it is a short-term effect
16 or whether it is going to last. We only have four
17 months of post-GDPR data. So that is something to
18 keep in mind. This is at least the short term that we
19 observe -- the short-term effect that we observe.
20 So in terms of robustness, we looked at the
21 pre-periods before May 25th. At least at the deal
22 level, the number of deals, we did not see an effect
23 before May. At the total dollar amount raised per
24 week, we do see an effect that starts a little bit
25 earlier. It starts in April, late April, kind of

1 crossing over to May, and we see it kicking in in
2 early May, really kicking in. So firms were reacting.
3 They were reacting to those announcements.
4 So as a robustness, we exclude May from our
5 sample and all the results still go through. As an
6 additional robustness, we exclude the period between
7 summer 2017 and summer 2018 to control for
8 seasonality, and the results still go through.
9 We top coded observations to reduce the
10 influence of outliers, of those huge deals, and the
11 results still go through. We categorize industries in
12 an unsupervised manner using techniques like K-means
13 or other machine-learning techniques, and the results
14 still go through. And we used other specifications,
15 and the results still go through. So we tried to
16 break the results and they do not break easily.
17 So what can we do with this? Well, our data
18 set, as I mentioned earlier, has some information
19 about employment numbers, employment ranges, how many
20 employees are employed per firm. And, obviously, we
21 see these dollar amounts decrease in deals, but what
22 does it say about welfare? It does not say much. We
23 cannot draw a welfare implication for this. It could
24 very well be that those less desirable firms are not
25 coming to fruition. We are preventing the next

1 Cambridge Analytica. Who knows.
2 But we can look at the effect on jobs. So
3 to do that, we got an average for the dollar amount
4 raised per employee by zero to three-year-old firms,
5 and that range is from $123,000 to a million dollars.
6 And we can use this range to see how many jobs are
7 lost because of the less dollars that come into those
8 firms. The fewer dollars that come in terms of the
9 number of investment deals and the dollars per deal.
10 Just if you are curious, how many dollars
11 are raised on average for a broader swath of firms,
12 say, zero to six-year-old firms, you see that those
13 dollars shrink. And the reason they shrink
14 potentially is because those firms have outside
15 revenue sources. I mean, they have their own revenue
16 sources, perhaps.
17 So those zero to three-year-old firms are
18 the most susceptible to job losses. They depend on
19 that money in order to hire those people. They depend
20 on those deals coming through in order to operate. So
21 in terms of jobs lost by those firms, based on our
22 back of the envelope, these rough estimates, we see
23 that it is between 3,600 and 30,000 jobs and that
24 amounts to about 4 to 11 percent of the number of
25 employees they employ in our sample.

1 I want to emphasize that this is the effect
2 we see in the short term. We do not know what is
3 going to happen in the long term. And it could very
4 well be that investors are just pulling out and saying
5 I want to see how this is going to shake up. I will
6 come back later. It could also be that investors are
7 shifting their dollars to the U.S., in which case, our
8 results may be overstated. It could also be that
9 there are investors outside the EU that tend to invest
10 in EU firms that hold their dollars back. We do not
11 see them in our sample because we only focus on the EU
12 and U.S., and so maybe our results are understated.
13 And the other thing to keep in mind is that
14 these jobs lost are just technology jobs in those zero
15 to three-year-old ventures, at least these rough
16 estimates. There could be more jobs lost. There
17 could be jobs lost by firms that are older. There
18 could be jobs lost by people who would have acted in
19 service positions for these jobs, providing lunch,
20 providing child care, and so forth.
21 So just to kind of summarize what we see so
22 far is that in the short run, we notice a pronounced
23 negative effect on EU venture financing, both on the
24 number of deals and the dollar amount per deal. Our
25 sample of post-GDPR is relatively short, so more study

1 is definitely needed here. And the reason that
2 investors are holding money back is not crystal clear.
3 It could be a wait-and-see approach. It could be that
4 they are afraid about rising compliance costs. It
5 could be that this regulation is hindering the actual
6 business practices that they want to invest in or the
7 products they want to invest in. It could just be
8 uncertainty.
9 The other thing to keep in mind is that our
10 sample is a small part of the bigger picture. We do
11 not have a complete universe. We think it is a pretty
12 good sample, but there could, of course, be more.
13 The other thing we notice here is that GDPR
14 is very transformative. It applies across categories,
15 even those categories we would expect may be less of
16 an effect because of existing laws like HIPAA.
17 So just one difference between HIPAA and
18 GDPR, one of many, is that HIPAA might require you to
19 provide consent in order to receive service from a
20 healthcare professional, whereas GDPR requires the
21 firm to provide service even if you do not give
22 consent.
23 In terms of Gramm-Leach-Bliley in financial
24 markets, that regulation provides an opt-out approach.
25 It basically allows customers to opt out of having

1 their data, say, sold to affiliates. Even that is in
2 special circumstances. Whereas GDPR requires an
3 opt-in approach, you have to provide opt-in consent
4 for your data to be used, for your data to be sold.
5 The penalties are also very different. GDPR
6 has much larger penalties, potentially 4 percent of
7 global revenues.
8 So aside from the negative effects we see on
9 the number of deals, we also have some conclusions or
10 at least preliminary conclusions for job losses, and,
11 again, it is a rough calculation. Other than that, I
12 would be happy to take some questions.
13 MR. STIVERS: Thank you, Liad.
14 So first of all, I would like to say to all
15 of you, hopefully a number of you are researchers in
16 this area, this is the kind of work that is incredibly
17 valuable, both to the FTC and to our sister agencies
18 working in this area, in terms of really trying to
19 understand what the potential effects might be of
20 changing regulation, changing the course in this area.
21 So if you are in this field, I strongly encourage you
22 to -- ah, there we go. I thought I had gotten the
23 button. I guess I had not.
24 Hopefully, you heard me that I strongly
25 encourage you to do research in this area. I know

1 that a couple of you have some very interesting work
2 coming forward in this area, so we are all eagerly
3 awaiting that.
4 However, since Liad is here, I get to grill
5 him a little bit. I wonder a little bit about the
6 time period that you are looking at. You look at the
7 time in which GDPR was actually -- the enforcement
8 happened. Did you think about looking at the April
9 2016 shift? Because you would expect that investors
10 maybe would be -- this was not a surprise, that it was
11 coming, even though I think you point out that perhaps
12 some of the companies were kind of last minute in
13 terms of getting their compliance up and running.
14 So can you talk a little about why you would
15 not necessarily see most of the effect happening right
16 around April of 2016, before and after, and then what
17 are you actually measuring? Are you measuring the
18 entire effect of GDPR when you look at the May 2018
19 date or is there something a little more subtle about
20 what you are measuring there in terms of the effect?
21 MR. WAGMAN: Right. So first, I would like
22 to say that I think both time dates are meaningful.
23 April 2016 is when GDPR passed, came into law, but it
24 was not to kick in until two years later.
25 Now, the second time period is meaningful

1 because that is the actual implementation stage. A
2 lot of these smaller firms that we focus on, they
3 depend on the policies that are adopted by the larger
4 firms, and those policies were not announced or not
5 adopted until the very few weeks, if not the week of,
6 May 25th, 2018.
7 So a lot of the realization of those
8 increased compliance costs, those increased liability
9 costs, the actual code that you needed to put in your
10 app in order to be compliant with the app store where
11 your app is published was not available until those
12 few weeks preceding May 2018, at least for the most
13 part. Just to give an extreme example, Google
14 released some code the day before.
15 Now, we looked at April 2016; in fact, we
16 started with that, and just our early checks did not
17 reveal a significant effect. It could be just, you
18 know, lack of clarity about what was going to happen.
19 Now, we saw that lack of clarity from regulators as
20 well. If you look at their own models for kind of
21 trying to predict what would happen after the
22 regulation, they had their own uncertainties. And
23 those uncertainties, I believe, are still not clear.
24 Until several probably lawsuits settle down, we will
25 not know the full effect.

1 MR. STIVERS: Okay. Thank you very much,
2 Liad. If you can thank our speaker.
3 MR. WAGMAN: Thank you.
4 (Applause.)
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

1
2 BIG DATA FAILS: RECENT RESEARCH INTO THE SURPRISING
3 INEFFECTIVENESS OF BLACK-BOX AI
4 MR. STIVERS: All right. We are going to
5 move to a recorded presentation from Catherine Tucker
6 of MIT. And as soon as we move forward in the slides,
7 it is going to start, which is why we still have
8 Liad's last slide up here.
9 Good, all right.
10 RECORDING: Good afternoon. My name is
11 Catherine Tucker and I am a Professor at the MIT Sloan
12 School of Management. Today, I am going to be
13 presenting some research I have into the surprising
14 area of big data in the online advertising world.
15 Before I start, I have two apologies. The
16 first one is obvious, I apologize very much for not
17 being at the hearings in person. I have teaching
18 scheduled on every single day of the hearings from
19 morning to afternoon, and I am very sorry not to be
20 with you. It looks like an amazing program.
21 The second apology is, unlike many of the
22 presentations you are going to see over these three
23 days, I am going to be presenting a research paper
24 today, and the nature of the research paper, of
25 course, especially an empirical one like this, is it

1 tends to go after a very narrow set of findings, but
2 makes sure that we can really believe in those narrow
3 set of findings. So the second apology is that what
4 you are going to hear is about a very specific set of
5 experiments in a very specific context.
6 So having said that, perhaps we should
7 actually move to the context. And as I alluded to,
8 this paper is a paper about big data in online
9 advertising. And to set the background, I want to
10 just remind you about how important data can be when
11 we are thinking about showing ads to a pair of
12 eyeballs on a particular website.
13 I also want to tell you about different
14 types of data that a publisher of the website --
15 imagine it's a news site and an advertiser could
16 potentially use to make sure they are showing the
17 right ads to the right person. The first thing they
18 could do is they could use something called first
19 party data. And that is data that the website
20 actually has access to because it knows the kind of
21 content that the consumer has browsed at some point in
22 the past. So if that news website knows that whenever
23 I see a cruise story, I read it, then perhaps they
24 could use that data to make sure they show me an ad
25 for an upcoming cruise.

1 Now, second party data is a little bit more
2 of a narrow pedigree and this is a capture view of
3 data where a website has data from a partner and they
4 know exactly who that partner is and what kind of data
5 they are getting. So a good example of that I came
6 across recently is that Rough Guides, a kind of travel
7 book, shares data, browsing data explicitly, with
8 lastminute.com, which is a travel website.
9 And you can imagine why they share data and
10 why it might be useful to working out what ad to show.
11 If someone has just booked a cruise to Italy, then if
12 I am Rough Guides, I want to show them an ad about my
13 guidebook to Italy, and similarly, if I am
14 lastminute.com and I found out that someone has been
15 buying guidebooks about Italy, it might be time to get
16 those Italy hotel ads up on my website. The key
17 thing, though, about this kind of data is that this is
18 data where everyone knows what it is and where it is
19 coming from.
20 The last kind of data -- and this is the
21 data I am going to be focusing on in this presentation
22 -- is something called third-party data. And this is
23 data purchased from a third-party source with the aim
24 of identifying what we call in marketing a customer
25 segment or a particular kind of customer you might

1 think is interested in your product or service.
2 Now, the actual purchase of this data is
3 extremely complex and is subject to a lot of different
4 technologies. I am going to simplify the terminology
5 slightly in this presentation and just talk about data
6 brokers. And you can think of data brokers as being
7 analogous to a data aggregator that comes and collects
8 all the different data sources from browsing behavior
9 across the web -- sometimes offline behavior, too --
10 into a file which summarizes all of the information
11 that is learned about a particular cookie or a
12 particular pair of eyeballs that is browsing the
13 internet.
14 Now, as you can imagine, these data brokers
15 have a lot of data. And as aggregate data, just
16 getting the pure data in place does not actually help
17 that much. You need to make inferences about who the
18 customer is and what they might be interested in if
19 you want to determine what ad to show them. And this
20 paper is going to be all about how good the algorithms
21 are which use this data to try and make inferences
22 about consumers and what kind of ads they might be
23 interested in.
24 So to give you an example of what I mean
25 about this third-party data, I thought we would start

1 with a specific example, and I am going to show you
2 how to do this with Twitter. Now, why Twitter? Well,
3 simply because it is actually quite straightforward to
4 get access to this kind of data on the Twitter
5 platform about yourself, and also my gut feeling about
6 the audience of the FTC hearing is many, many people
7 have a Twitter account.
8 So what you should be doing right now is
9 getting out your mobile if you are not already playing
10 around with it and follow along to see how you can
11 find out what data Twitter has about you, which is
12 this kind of third-party data where people or
13 algorithms have made inferences about your profile as
14 a consumer.
15 So what you do is you get out your Twitter
16 profile and you go and look at settings and privacy.
17 You can see that I have highlighted it right on the
18 left-hand screen right there. And then after that,
19 you go and choose -- you go to the privacy and safety
20 screen and you scroll down to the bottom where you
21 have the opportunity to see your Twitter data.
22 Then on the next screen, I would like you to
23 select the second option, which is this third-party
24 data, which is all about inferred interests that
25 Twitter has from third parties who have been

1 collecting data about your browsing of the internet.
2 Now, if you click on this with me, I will
3 show you what I see. So you are going to see a whole
4 lot of different things that this third-party data and
5 the algorithms have inferred about you. This is what
6 they have inferred about me.
7 Now, here you can see that they think I have
8 one child. I actually apologize to my other three
9 children, I obviously do not browse enough about you.
10 You can also see that my web-browsing patterns has led
11 Twitter has inferred that I am actually a senior in
12 terms of my age range.
13 I think probably the thing I worry the most
14 about is how it is that these third-party brokers have
15 inferred that I am a single parent. I think, at this
16 point, I really do have to apologize to my poor
17 husband.
18 Anyway, the key thing here for the purposes
19 of this talk is that you can see demographics, what
20 they have inferred about your demographics, right,
21 because, in general, a pair of eyeballs browsing on a
22 mobile phone or a desktop, there is no real way of
23 sort of telling, you know, exactly what your
24 background demographics are. So the algorithm is then
25 going to use the data about your browsing to try and

1 infer what your demographics are from your browsing,
2 and that is going to be the focus of this study.
3 Now, the specific name of the paper I am
4 going to be talking about, if you want to read it in
5 details and, you know, go into all the nitty-gritty,
6 it is up on SSRS and you can easily find it there and
7 it is called "How Effective is Black-Box Digital
8 Consumer Profiling and Audience Delivery?: Evidence
9 from Field Studies."
10 I should highlight that this is not work I
11 have done by myself. Instead, I have a wonderful team
12 of coauthors. Nico Neumann is at the University of
13 Melbourne Business School and he is an amazing very
14 junior professor who really cares about this industry
15 and trying to work out what is going on, and Tim
16 Whitfield, who was actually at one of the large
17 advertising agencies at the time we wrote the paper,
18 and he organized for us to get access to this world to
19 study how well it works. So I owe a huge gratitude to
20 my coauthors.
21 This paper consists of three separate
22 studies, and in all these studies, we are asking, how
23 well does the big data and online advertising
24 ecosystem do in terms of identifying gender and age.
25 Why gender and age? Well, first and most importantly

1 for us, they are things you can actually potentially
2 verify.
3 The second reason -- and this is actually a
4 very popular form of data that advertisers use for
5 targeting the -- if you look at it from industry
6 surveys at least, the age data, gender data tend to be
7 most broadly used types of data for the targeting of
8 ads.
9 Now, the way we proceeded, as I said, there
10 were three studies and in each study, we actually
11 tried to make the task of identifying whether a
12 particular pair of eyeballs was from a certain gender
13 or a certain age easier and easier. The first study
14 was the most broad-brushed, and as such, I will go
15 through it quickly.
16 And what we did there was we went to various
17 ad platforms and said, can you show our ad 100,000
18 times to men between the age of 25 and 54. When we
19 gave them this simple mission, there was a large range
20 of success, but we found they were able to do this, on
21 average, about 59 percent of the time when we compared
22 their performance with our benchmark, which was the
23 Nielsen data that actually reported the age and gender
24 of the eyeballs that were seeing our ads.
25 Now, in some sense, to be clear, this is an

1 improvement relative to sheer chance. Sheer chance
2 would be below a third given the makeup of the
3 internet compilations. There is an improvement of 184
4 percent when we use the data ecosystem to try and
5 enhance our advertising. I think, though, the point
6 we are trying to make in the paper is, yes, there is
7 definitely an improvement. But given that advertisers
8 tend to be paying more than -- 200 percent more to
9 show their ads using these data-targeting tools rather
10 than just showing them by chance to everyone, it was
11 not quite clear to us that the return on investment
12 was there.
13 Now, as our first study -- and you might say
14 this is somewhat unfair because it was still relying a
15 lot on humans to have discretionary choices perhaps
16 about how they set up the campaign and that could
17 explain the failure we are seeing. So in our next
18 study, we wanted to try and take out that human
19 element.
20 What we did for the next study was we tried
21 to make it easier for data brokers to do this. So we
22 sort of tried to take out the human element. And so
23 in our second study, what we did is we said we have
24 this website, please, data brokers, tell us who the
25 audience of the website is. So there was no

1 discretion in finding particular eyeballs; you just
2 have to tell us who the eyeballs at a website is.
3 Now, when we did this, we did this test with
4 four separate data brokers. On average, what was just
5 striking is that they told us in terms of proportion
6 of men it is 58 percent, it is 55 percent, 85 percent,
7 63 percent. I am not sure what we can say about
8 accuracy here. It does not seem great to me. If I
9 got back those numbers, I would still not quite know
10 what the true proportion of men is.
11 What was also striking to me about this
12 study, and we should see it in the paper, is that
13 never mind getting their gender right, they had no
14 idea when we asked people what the actual number of
15 eyeballs was on these websites. At least those
16 huge -- when I say “no idea” what I mean is there is
17 huge variation in the answers we were given, which
18 ranged all the way from 300,000 to 500,000 eyeballs,
19 which is a large difference if you are an advertiser.
20 So the second study did not give us much
21 reassurance that we were really getting accurate
22 information here. So what we decided to do in our
23 third study was to just make the task as simple as you
24 could ever possibly imagine. And in this task, what
25 we said to each data broker was, look, you do not have

1 to tell us about a particular website, all you have to
2 do is tell us do you have data or a profile about this
3 particular cookie, and if you do, can you tell us what
4 gender you think this cookie and the set of eyeballs
5 associated with this cookie are.
6 Now, you might be saying, okay, you keep on
7 saying we know really how many -- you know, what
8 gender people are, how do you know the truth, and what
9 we did in this study to find out the truth, which, you
10 know, I find quite compelling, is that we used a
11 service named Pureprofile to actually verify what the
12 truth is. And what Pureprofile goes out to do is they
13 actually survey people, and so they go out and say,
14 what gender are you, what age are you, and they give
15 you a [indiscernible]. So we used that as our source
16 of truth about what the true gender and true age is.
17 And you may, of course, be cynical and say,
18 well, are all people in an online survey really going
19 to be completely honest? And, of course, I am sure
20 there are some people who are not honest when asking
21 these surveys. However, it is said to be our source
22 of truth and at least it is what we can call declared
23 data for what people want me to think about their age
24 and gender. So we are going to use that as our
25 measure of the truth. And the question is, what did

1 we find when we compared this declared data to what
2 the data brokers were telling us.
3 And you can see here we actually used a lot
4 of different data brokers in this study, and there was
5 a wide range of how many cookies they told us they had
6 information for, and you can see that in the second
7 column.
8 What I want you to look at, though, is the
9 third column. And we actually asked them the specific
10 task of telling us whether or not that cookie was
11 male. And that is going to be our measure of gender
12 accuracy. And the number you see in the third column
13 is the percentage of times they were able to correctly
14 tell us that a cookie was male.
15 And I want you to look at those numbers and
16 also register the fact which I always found the most
17 hilarious about this study and this paper in general
18 is that it is if you sort of take the average of
19 accuracy really pretty close to 50 percent -- in other
20 words, these data brokers, this entire big data
21 ecosystem, seem to be able to tell us the gender of
22 the pair of eyeballs correctly half of the time. And
23 if you have ever taken probability theory and you have
24 thought about the distribution of men and women, you
25 will see why this is quite funny.

1 The other thing I want you to look at in
2 this table is the second column, which is the number
3 of cookies. The reason I think this is important for
4 this meeting -- I do not usually emphasize it when
5 presenting the paper, but I think it is interesting --
6 is as a subset you might think of this as a measure of
7 how much data the data broker is really working,
8 right. Because we asked them, well, how many cookies
9 can you tell us about and so it seems reasonable to
10 infer that if they could tell us about more cookies,
11 they have more data.
12 Now, the reason this is important is that if
13 you were to try and think about a correlation between
14 the second column and the third column, and look to
15 see is there any relationship between the amount of
16 data these data brokers appear to have access to and
17 how good they are at telling the gender correctly, you
18 know, there is not really enough data points to run a
19 regression, but I [indiscernible] just see no
20 available correlation really whatsoever. So I think
21 it is important because it suggests that there is a
22 surprising lack of correlation between access to data
23 and how well these data brokers are performing in
24 terms of being able to use an algorithm to infer
25 gender.

1 So let's just summarize the findings of
2 this research. So back in the big headline news --
3 and this is going to spill over, I'm sure, into the
4 meetings in two weeks time -- is that, in general, we
5 have often worried about algorithms, big data, AI, and
6 we tend to worry though more from an Orwellian privacy
7 intrusive way. However, I am here to tell you we
8 might be worried about these algorithms being too
9 accurate, but I am really worried about the fact that
10 they seem to be surprisingly bad at actually getting
11 something very basic like being able to infer gender
12 from browsing behavior.
13 Now, it seems very straightforward that when
14 you think about it, maybe there is a reason these
15 algorithms are doing not bad, but poorly. I mean, I
16 challenge everyone in this room to think about the
17 internet sites you browse and really how informative
18 are they about gender? I can imagine that there are
19 perhaps some particular websites which tell you a lot
20 about gender, maybe a website devoted to the merits of
21 sanitary products or something like that. I do not
22 think there are probably many men browsing those types
23 of websites.
24 But, in general, if you think about the
25 right browsing behavior, it is talking about gender

1 and I think that is just an overarching problem these
2 algorithms are facing. They are trying to infer
3 something which maybe is just not inferable given how
4 different our -- “browsing behavior” given how
5 different genders really perform -- use the -- how
6 people with the same gender use the internet.
7 The other reason this is going on is
8 actually be even more simple and, you know, this is
9 not a complete explanation, but it is certainly a
10 partial information, but one of the reasons these
11 algorithms appear to be failing is that we looked to
12 see how does that accuracy vary with household size.
13 And we showed that as your household gets bigger and
14 as you have more than one person potentially using a
15 computer or a device, then the accuracy does appear to
16 fall.
17 So a simple explanation, we are trying to
18 infer gender potentially from a computer, which in my
19 case is used by my husband, used by me, used by my
20 kids to watch My Little Pony videos. It is going to
21 be very hard to actually work out what gender a pair
22 of eyeballs are when you do not have just one pair of
23 eyeballs.
24 Now, another point I want to make is not
25 just that this kind of data inference process in the

1 use of algorithms on big data does not seem to provide
2 necessarily insights that we might fear it does in
3 terms of how accurate it is, it is just because these
4 are hearings about competition is that you often hear
5 repeated the mantra, the idea that there is a link
6 between access to data and the ability to compete.
7 And especially in a world of algorithms, you
8 can see the argument for that and that perhaps if I
9 have a larger data set, I can train my algorithm to
10 perform that much better and be able to outcompete my
11 rivals. However, what I saw in this study, at least
12 in this early -- potentially early and nascent stage
13 in this industry is that the size of data did not seem
14 to matter that much, or really at all that I could see
15 in the data, of how well these data brokers were doing
16 in terms of accuracy.
17 And that suggests perhaps an argument which
18 I think we will probably be hearing about in two
19 weeks, that really the quality of algorithms are going
20 to be potentially more important than the quality of
21 [indiscernible] these algorithms may end up being more
22 important than the actual size of data that are used
23 to train these algorithms.
24 So with that, I will say thank you so much
25 for listening. Apologies again for not being in

1 attendance. It does look wonderful. And if you have
2 any questions, feel free to email me. Thank you very,
3 very much.
4 MR. STIVERS: Thank you, Catherine, in
5 absentia.
6 (Applause.)
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

1
2 CORPORATE DATA ETHICS: RISK MANAGEMENT FOR THE
3 BIG DATA ECONOMY
4 MR. STIVERS: All right. Our next speaker
5 is Dennis Hirsch from The Ohio State University Moritz
6 School of Law.
7 MR. HIRSCH: Commissioners and FTC staff,
8 thank you for inviting me here today and giving me the
9 opportunity to present my research at this hearing. I
10 am going to discuss one of the hearing's principal
11 topics, whether companies can use improved privacy
12 performance for competitive advantage.
13 To address this topic, I need first to
14 slightly reframe it. The question should not just be
15 whether companies can use improved privacy performance
16 to achieve competitive advantage, but whether they can
17 use more responsible data practices at large to do so,
18 including issues of bias, procedural fairness, and
19 manipulation. Some companies are doing this, and they
20 have a name for this broader project. They call it
21 data ethics.
22 I am currently leading an Ohio State
23 research project that is studying corporate data
24 ethics and, today, I am going to share with you the
25 preliminary findings from this research and I will

1 address four questions. One, what is data ethics?
2 Two, why are companies engaging in it? Three, how are
3 companies trying to achieve it? And four, what does
4 this mean for the FTC's regulatory authorities?
5 I was led to this topic a couple of years
6 ago when at a roundtable discussion I heard the chief
7 privacy officer for a large company say that her
8 department was debating what was ethical to do with
9 data and what was not ethical to do with it. And this
10 surprised me. She was a chief privacy officer, why
11 wasn't she worrying about compliance with privacy
12 laws?
13 And when I began to hear about other
14 companies engaged in the same activity, I thought it
15 would be interesting to study this phenomenon. I put
16 together a terrific team of colleagues, faculty
17 colleagues from the schools of business and computer
18 science, philosophy and sociology, and together, we
19 decided to use three methods to try to address this
20 question, a literature review, expert interviews, and
21 a broad survey of companies that use big data
22 analytics.
23 So today, I am going to present our
24 preliminary findings, but first I need to make two
25 caveats. One, we have completed the literature review

1 and we are midway through the interviews, but we have
2 not done our survey as of yet. So this truly is
3 preliminary findings. We are still in the midst of
4 this study.
5 Second, our interviews focus on corporate
6 managers at large companies. So we are not getting a
7 comprehensive view of Corporate America, nor
8 necessarily are we getting a fully objective view.
9 That said, I think we have been getting some valuable
10 information that I will try and share with you today.
11 So as told to us by those that we
12 interviewed, the story starts with big data analytics
13 and its sister technologies, machine learning and
14 artificial intelligence. Now, it is well-known that
15 these technologies can create many benefits, some of
16 which we have heard about already today. But what the
17 companies told us is that they also produce important
18 risks.
19 And they identified four types of risks:
20 Risks of privacy violation, such as when Target used
21 predictive analytics to infer from customer purchasing
22 histories whether its female customers were pregnant;
23 risks of bias, as when Amazon recently discovered that
24 the artificial intelligence application it hoped to
25 use to sort through the thousands of resumes that it

1 received was systematically discriminating against
2 women, and Amazon caught that problem and decided not
3 to use that AI application; risks of procedural
4 unfairness as when black-box algorithms, which are not
5 subject to explanation or appeal, are used to inform
6 decisions whether to grant loans or jobs or housing;
7 and risks of exploitation or manipulation such as when
8 Cambridge Analytica used Facebook users' data to infer
9 the psychological types of those users and target them
10 with political ads that they would find hard to
11 resist.
12 As the companies see it, these potential
13 harms threaten not just the individuals in question,
14 but also the reputation of the companies themselves,
15 and this creates an urgent issue for these companies,
16 which is how to reduce these risks. As one corporate
17 manager put it to us, if data use has much more
18 impact, then you need a governance structure to help
19 manage the impact of that data use to make sure the
20 organization does not create more risk for itself.
21 Now, traditionally, companies have mitigated
22 digital risk by complying with privacy laws, but --
23 and this is a key point -- big data analytics renders
24 that insufficient. And it does so for two main
25 reasons. First of all, the risks that I just

1 mentioned start with privacy, but they go well beyond
2 it to bias, procedural unfairness, and manipulation.
3 So privacy law is not going to be sufficient to
4 address that.
5 Second, privacy law is premised on the idea
6 that given accurate notice, individuals can make
7 choices about what companies can do with their data.
8 So by making such choices, individuals can protect
9 themselves. But big data analytics changes this. It
10 allows companies to take surface data and infer latent
11 information from it. For example, it allows Target to
12 take customer purchasing histories of its female
13 customers and infer whether they are pregnant.
14 Given this ability to infer latent data from
15 surface information, people cannot know what they are
16 really revealing when they decide to hand over the
17 surface information. And as a result, they cannot use
18 notice and choice to protect themselves, at least when
19 it comes to big data analytics, machine learning and
20 AI. From the company's perspective, this means that
21 if they are going to protect individuals against the
22 risks that these technologies pose and so protect
23 their own reputations, they have to do more than
24 comply with privacy law. They have to ensure that
25 their practices are also ethical.

1 So here is what one lawyer who advises such
2 companies said to us: Preying on vulnerable
3 populations, treating people unfairly, manipulating
4 people in ways that could harm them, there is some of
5 that stuff that is perfectly legal, but it still may
6 not be a good business decision. I will throw out the
7 word "ethics." It is not the ethical thing to do.
8 Some companies that I work with, they take that stuff
9 very, very seriously. They do not want to do things
10 that feel or could be perceived as unethical.
11 Now, some, including some of our colleagues
12 in Europe, see data ethics as an attempt to take
13 Kantian or Aristotelian or other ethical philosophies
14 and use them to govern advanced data practices. But
15 that is not what we saw these companies doing. For
16 them, data ethics is beyond-compliance risk mitigation
17 for the big data economy. Hence, the title of my talk
18 today.
19 So that is what data ethics is. Why do
20 companies seek to achieve it when existing law does
21 not require them do so? We identified three principal
22 motivations: Reputation, employee retention, and the
23 threat of regulation. I have already mentioned
24 reputation, but the companies tell a more nuanced
25 story about it. There is reputation among customers

1 and users. This is essential to preserve the bonds of
2 trust on which the flow of personal data depends. As
3 one company manager said to us, if you act ethically
4 and ensure that data use is ethical and you are fully
5 accountable for that, then your brand is trustworthy.
6 That is what we are all trying to achieve.
7 Then there is reputation among regulators
8 and advocates and a poor reputation among these
9 constituencies can lead to increased scrutiny in
10 litigation. And, finally -- and this is the one that
11 surprised us a bit -- there is reputation among your
12 business partners. A lawyer for one technology
13 company saw this as the most important aspect since
14 other businesses are able to do due diligence in ways
15 that individuals cannot and will not work with
16 companies that do not pass muster.
17 Employee retention was a third major driver.
18 Tech companies, in particular, expressed that
19 competition for young engineers is fierce and is
20 critical to corporate success and that companies need
21 to align their actions with these young people's
22 values in order retain them.
23 The third driver we saw was the threat of
24 regulation. Some companies believe that if they took
25 proactive steps to act responsibly, they would reduce

1 the chance of direct regulation, data ethics as a way
2 to preempt direct regulation. Others, with an eye on
3 the GDPR and other rules, saw data ethics not so much
4 as a way to avoid data regulation, but as a way to
5 prepare for it. They felt that if they aligned their
6 products and systems in advance, they would be able to
7 deal with such regulations more effectively and at
8 less cost than their competitors.
9 So with each of these drivers -- reputation,
10 employee retention, threat of regulation -- companies
11 are seeking a form of competitive advantage. And
12 thus, our research suggests that corporate data ethics
13 represents a new form of competition in the
14 algorithmic society, one that goes beyond just
15 competing on privacy attributes. One leading privacy
16 professional put it this way, “I think that for some
17 of these companies, they have actually seen data
18 stewardship as a competitive differentiator and that
19 they are more trustworthy and people are more likely
20 to do business with them and, therefore, pay higher
21 prices.”
22 I should add that several interviewees
23 expressed that their company’s values were also very
24 important in driving their data ethics initiatives,
25 and that was particularly true where a CEO or a

1 founder had instilled those values particularly
2 strongly. So that can also be a motivator.
3 Now, we have looked at the what and the why
4 of data ethics. The next question is the how. Here,
5 it is helpful to divide this into two areas, process
6 and substance. In terms of process, one of the really
7 interesting developments that we found is the
8 transformation of the privacy officer role into a role
9 that included not only privacy, but also issues of
10 bias and procedural unfairness and manipulation.
11 Reflecting this, some companies changed the
12 title of the position to include the word “ethics” or
13 “data ethics” in it. This is a new development that
14 has just arisen, we think, within the last year. But
15 it could soon be common to have a chief data ethics
16 officer to go along with your CIO or your CISO or your
17 CPO.
18 Another interesting development was the
19 creation of new committees to advise the companies on
20 ethics. Some created internal committees, sometimes
21 called an ethics review committee, to review data
22 analytics projects that raised ethical risks. Such
23 committees could include representatives from legal,
24 privacy, security, engineering, and the affected
25 business unit, and we saw instances in which such a

1 committee advised against certain projects and the
2 companies turned down significant contracts on this
3 basis.
4 Other companies ran their ethical questions
5 by external committee, sometimes called external
6 advisory boards, that might include privacy and
7 consumer advocates or members of civil rights and
8 civil liberties groups or academics. In contrast to
9 the internal boards, these served in purely an
10 advisory role and helped to sensitize the company to
11 stakeholder concerns.
12 There was quite a bit of variety in the way
13 the companies managed in this area. For example, they
14 differed on the scope of their ethics management
15 activity. Some focused on the company's own internal
16 research with customers, personal information; others
17 expanded the scope to include not only their own
18 activities, but also those of data suppliers,
19 customers, and business partners, anyone whose ethical
20 lapses could be linked to them.
21 I practiced and taught environmental law
22 before I turned to data and privacy and these programs
23 reminded me of the way in which some companies audit
24 the environmental performance of their entire supply
25 chain, a process they call greening the supply chain.

1 Companies also diverged in terms of
2 management structures and reporting systems. Some
3 localize the ethics function in a single person who
4 had a direct line of communication to the C-suite or
5 CEO. Others had a far more elaborate process in which
6 all data projects had to be submitted for review. As
7 we understood it, the first seemed to produce faster
8 decisions; the second, better quality decisions. So
9 there is a tradeoff here.
10 Turning from process to substance, we sought
11 to identify the standards that companies employ to
12 assess whether a given data analytics project is
13 ethical or not. The literature suggests that
14 companies employ or should employ formal principles
15 grounded in philosophies of ethics. For example, the
16 Software and Information Industry Association that
17 Mark MacCarthy works with -- and he was here today --
18 they draw on such ethical traditions and published a
19 report that articulated four core principles --
20 rights, justice, welfare, and virtue -- that companies
21 should follow when making decisions about ethics.
22 The companies we talked to were not using
23 any such formal framework. What we saw was far more
24 intuitive. One manager referred to the quote
25 “fairness check,” which the manager described as would

1 my mother think this is okay. Would I want this to
2 happen to my kid? Do I feel good about this
3 personally?
4 Another employs the ear test. Saying the
5 ear test simply means to me, does that sound right?
6 Does that sound like a bad idea? Do the words coming
7 out of your mouth make sense from both a legal,
8 ethical and business standpoint?
9 So these companies are using much more
10 intuitive expectation-based standards rather than
11 formal philosophical ones. Such standards fit with
12 the idea, mentioned earlier, that companies are
13 seeking not to implement an ethical philosophy, but
14 rather to engage in beyond-compliance risk mitigation.
15 In this sense, data ethics is a new dimension of
16 corporate social responsibility. It is CSR for the
17 data-driven business.
18 Responsibility, appropriateness,
19 trustworthiness, fairness, these seem to be the
20 currency of data ethics. Now, these can be difficult
21 concepts to operationalize and some companies seem to
22 really struggle with drawing these lines. The hardest
23 question seems it to be how to get the balance right,
24 how to determine, considering the potential benefits
25 and risks, what is fair and what is not. As one

1 attorney said to us, when do these lines get crossed?
2 That is not always obvious.
3 What does all this mean for a regulator,
4 like the FTC? Well, when you step back from what we
5 have learned so far, you really see two things. You
6 see a pretty clear consensus among the larger, more
7 sophisticated companies, at least, that it is
8 important to go beyond compliance and seek to mitigate
9 the risks that big data analytics can pose.
10 So there is quite a bit of agreement on the
11 what and the why. But the how question is much more
12 murky. Companies are experimenting with many
13 management processes and trying to figure out which
14 will be more effective, and there is some confusion as
15 to how to draw the line between responsible and
16 irresponsible behaviors.
17 I mentioned that I came to privacy from the
18 environmental field. And this situation reminds me in
19 some ways of that which environmental regulators faced
20 when companies started to compete seriously in terms
21 of their environmental performance, which is known as
22 green business. One thing that environmental
23 regulators did and that the FTC could do is to collect
24 and share best practices in this area as a way of
25 getting more companies to adopt them.

1 Another would be to adopt a leadership
2 program that recognizes companies that are going above
3 and beyond in this area and so add to the reputational
4 value they derive from doing so.
5 A third would be to define some standards in
6 this area. Now, I would caution against doing this
7 with respect to process. There seems to be a lot of
8 positive experimentation going on and regulators may
9 want to let that play out before determining that one
10 approach is preferable to another, but it may be worth
11 giving this further thought with respect to drawing
12 the substantive lines.
13 Were a regulator to provide some guidance,
14 that could give companies a clearer sense of what the
15 regulator's expectations are and help them to make
16 some of the tough calls. It could also set a floor
17 that all companies have to pay attention to. Right
18 now, we are seeing the larger, more sophisticated
19 companies start to manage data ethics. But other
20 companies that are not paying attention to these
21 issues could do some really bad things that could not
22 only hurt people, but could also turn the public
23 against data analytics, machine learning and AI more
24 generally.
25 As one attorney said to us, "In this fast-

1 paced world where there is, you know, huge financial
2 opportunity for companies, you can easily see
3 scenarios where someone is going to, quite frankly,
4 bring down the whole house of cards by doing something
5 just totally unethical and totally unfair and screw it
6 for the rest of the industry." Seen in this light, a
7 regulator's decision to set a floor for fair and
8 ethical behavior could potentially support the efforts
9 of the current leaders while still giving them room to
10 distinguish themselves.
11 If the FTC wanted to develop such
12 substantive guidelines, rules of the road for
13 predictive analytics, it seems to me that it has the
14 power to do so. The line that companies are trying to
15 draw is between advanced analytics that is
16 appropriate, and that which is unappropriate, between
17 that which is responsible and that which is
18 irresponsible, between that which is fair and that
19 which is unfair.
20 Section 5 of the FTC Act, of course, gives
21 the Commission the power to define unfair business
22 acts or practices and so to draw these lines. The
23 FTC's unfairness authority has some useful features in
24 this regard. Unfairness is an open and flexible
25 standard intended to adapt to emerging and changing

1 technologies and business models. It requires the
2 Commission to balance benefits and costs, which is
3 important in an area like big data analytics that
4 offers many benefits, as well as many risks. And
5 unfairness is intended precisely for those situations
6 in which individuals cannot protect themselves, where
7 in the language of Section 5(n), the injuries are not
8 reasonably avoidable by the consumers themselves.
9 That is where we are with respect to
10 advanced analytics. Individuals cannot understand how
11 these technologies work and so cannot use traditional
12 privacy protections, notice and choice, to protect
13 themselves. Some companies are moving proactively to
14 protect them. The FTC could potentially use its own
15 unfairness authority to support this corporate data
16 ethics effort.
17 Thank you for letting me share my thoughts
18 and research with you today.
19 (Applause.)
20
21
22
23
24
25

1
2 FREE SPEECH AND DATA PRIVACY
3 MR. STIVERS: Next, we have Jane Bambauer
4 from Arizona, Rogers College of Law. Here we go.
5 Thank you.
6 MS. BAMBAUER: Thanks for having me. I want
7 to do something that will give me a little information
8 and give you a chance to wake up. So please stand up
9 if you can, if you are willing and able, and then if
10 you are not a lawyer, sit back down. I just want to
11 get a sense of what this audience -- if you are not a
12 lawyer -- okay, so that is about half of you.
13 And then, for those who are still standing,
14 if you do not know what the case Sorrell vs. IMS is
15 about, sit down. Oh, good. Okay, this is going to be
16 valuable. Thank you.
17 Okay, all right. So my goal with this talk
18 is to provide a descriptive account of what is going
19 on in First Amendment law and the ways that it might
20 actually limit some of what the FTC may want to do,
21 even if it has set its sights on what it thinks is the
22 best policy. I am going to be as descriptive as I can
23 without letting my own policy preferences kind of
24 shape that description because I will have a chance
25 tomorrow to talk again about what I think about best

1 policies. So I am going to do my best just to be sort
2 of honest about what is going on in the courts. But
3 the courts are being quite active in this sphere and
4 you all should pay attention.
5 Okay. So before I dive in, though, let me
6 tell you a little bit about what I am not going to
7 talk about. I will not talk about restrictions on
8 commercial speech. The commercial speech doctrine, it
9 is a little bit misleading. It is actually a narrow
10 category that covers just marketing messages and
11 advertising. So some people sometimes think that
12 commercial speech, which gives lesser protection to
13 commercial -- it is a doctrine that gives lesser
14 protection, that that would apply any time someone is
15 selling something to someone else or any time that
16 somebody has a commercially-motivated purpose to say
17 something and that is not accurate. Like books are
18 sold, right, and obviously books receive full
19 protection.
20 So commercial speech doctrine is narrow and
21 it is related to potential privacy regulation because
22 privacy laws often have as at least one of their end
23 goals to affect how marketers can craft messages. So
24 in that sense, it may be related.
25 Also, if companies are giving false

1 assurances, either explicitly or implicitly, to
2 consumers that their privacy is protected when it is
3 really not, that is related to the commercial speech
4 doctrine. The commercial speech doctrine does not
5 allow any protection, any First Amendment protection
6 to false and misleading commercial claims.
7 And that is interesting in what it means to
8 be misleading, especially when a company is committing
9 an omission, when they are not saying anything. It is
10 still an open question about whether that is
11 sufficiently misleading to remove the company from the
12 ambit of First Amendment protection, and there are
13 some papers, including one that I have written
14 recently, that kind of have tackled this question of
15 who gets to decide what it is to be false or
16 misleading -- and by the way, I am sorry, I mangled
17 Rebecca Tushnet's article. Hers is actually called
18 "It Depends on What the Meaning of False Is." I was
19 writing these slides late.
20 But we engage in a little bit of a debate
21 about who should decide and whether the courts need to
22 be involved. And it is interesting, but I am not
23 going to talk about it today.
24 The other interesting thing that is out of
25 scope for today is the compelled speech doctrine. So

1 that is related to privacy because regulators might be
2 interested in something like just-in-time privacy
3 disclosures that make clear notice about how a company
4 is going it use your data and we may be interested in
5 forcing companies to actually provide these
6 disclosures.
7 And the Supreme Court has given its blessing
8 to mandated disclosures that are purely factual and
9 uncontroversial information, so maybe like nutrition
10 labels on food items. I think most people tend to
11 think of that as purely factual. But it is not clear
12 whether a privacy policy would be -- or mandated
13 privacy policies would be purely factual and
14 uncontroversial. And I talk about this at some length
15 in another article. So if you are interested in this
16 topic, you can see my article that tries to map out
17 what courts, especially lower courts, have done to
18 decide whether a factual mandated disclosure is
19 ideological and, therefore, subject to constitutional
20 review or merely informational and not subject to any
21 amount of review.
22 Interesting stuff, but we do not have time
23 because I want to get straight to the core of what
24 almost every privacy law is going to wind up
25 potentially coming into conflict with, and that is a

1 restriction on noncommercial speech.
2 So this usually will happen in the course of
3 privacy regulation through one of two ways. Either a
4 law will put a limit on the transfer of personal data
5 between, say, one company and another or it will put a
6 limit on the initial collection or maybe even the
7 initial inference based on already-collected data
8 about a person. And, you know, almost every privacy
9 law, if you think about the FIPPs, the Fair
10 Information Practice Principles, they usually involve
11 giving the data subject some amount of control over
12 these two activities. And that control necessarily
13 puts a limit on these data transfers or the data
14 collection.
15 So much of what I am going to say, but not
16 all of it, is lifted from an earlier article I did
17 called “Is Data Speech?” asking, well, okay, is the
18 First Amendment relevant here? Do we need to worry
19 about potential constitutional review when we are
20 dealing with data privacy?
21 So let’s start with the -- oh, that is
22 right. To ground the discussion, I would like to
23 have you, in the back of your head, thinking about
24 the California Consumer Privacy Act because I think
25 that -- for many consumers, that seems to be a model

1 privacy law. It seems to tap into what many people
2 want or at least believe that they want.
3 And the most important rights that are
4 relevant for my discussion is that it gives
5 Californians the right to say no -- this is taken from
6 the website of the designers of the law -- it gives
7 Californians the right to say no to the sale of
8 personal information. It also, by the way, gives them
9 the right to demand the deletion of personal data
10 unless it is required for the service of the company.
11 And just like with the GDPR, if a
12 Californian does opt out of data sale, for example,
13 they still must be given service on the same terms as
14 somebody who has not opted out. But unlike the GDPR,
15 it is an opt-out regime rather than an opt-in regime.
16 Okay. So as I tell you about some of the
17 case law, work with this hypo -- law professors love
18 hypos, so ask yourself, okay, how does this affect the
19 constitutionality of California's recently adopted,
20 but not yet implemented, law?
21 Okay. I am going to start with data
22 transmissions. These little stick figures are meant
23 to be like companies or people who are selling data,
24 and that red thing is data.
25 So the first question that free speech

1 lawyers generally ask is, well, is the First Amendment
2 even relevant here? Does it cover this kind of
3 activity? Would we call this activity speech? And I
4 am starting with this rather than the initial data
5 collection, even though it usually comes later because
6 I think this question is actually much easier to
7 answer. I think courts are converging on a clear,
8 yes, this is speech, this is covered.
9 So the Supreme Court itself in earlier cases
10 had found that really dry information, like credit
11 reports or beer ingredients, are speech and really
12 anything that communicates from one person or entity
13 to another is speech. The lower courts, too, found
14 that even in the context of privacy laws that the
15 privacy laws may survive scrutiny, but that scrutiny
16 must be used.
17 Then the case of Sorrell vs. IMS, which most
18 of you do not know about which delights me because I
19 can tell you about it, really made this even more
20 clear. So this was a case from 2011 or 2012 involving
21 a Vermont statute that banned the sale of prescription
22 data to pharmaceutical companies if the pharmaceutical
23 company was going to use the data to fine-tune the
24 detailing, basically the marketing messages that it
25 made for doctors. So the data did not have the

1 identities of the patients, but the data does have
2 identities of the doctors. So you can see it was
3 justified partly on privacy grounds and partly on
4 public health grounds.
5 And as a privacy law, this seems rather
6 narrow, but you can see how the implications might
7 affect other types of broader privacy laws because if
8 you think of doctors as standing in for consumers
9 here, the law was trying to give doctors the
10 opportunity -- they could opt-in to these types of
11 marketing messages based on their data if they wanted
12 to, but it was trying to give them some control such
13 that behavioral advertisers basically would not have a
14 lot of detail about their habits.
15 So the Supreme Court -- by the way, some
16 commenters and even the circuit courts that were
17 hearing similar cases before Sorrell was decided
18 thought this type of law would fall outside the First
19 Amendment protection completely because data that is
20 just like sitting in a server and that is just sold
21 for these types of purposes is no different from any
22 other product. I think the First Circuit even said it
23 is like the equivalent of beef jerky -- selling beef
24 jerky. The Supreme Court definitely rejected that.
25 So in an opinion by Justice Kennedy, he

1 begins the analysis by saying this Court has held that
2 the creation and dissemination of information are
3 speech within the meaning of the First Amendment.
4 Facts, after all, are the beginning point of much of
5 the speech that is essential to advance human
6 knowledge and to conduct human affairs.
7 In the end, it got a little confusing
8 because the case was ultimately decided on grounds of
9 viewpoint discrimination because what at least Justice
10 Kennedy thought was the most -- the biggest offense
11 about this law was that it prevented only
12 pharmaceutical companies from using this type of tool
13 to craft their messages to try to persuade doctors to
14 do something, and it left open any other speaker who
15 was trying to persuade a doctor to do anything else,
16 it left access to the data open to them.
17 So the case was ultimately decided on
18 viewpoint discrimination grounds but the dicta that
19 came earlier seems pretty compelling and especially
20 because it is consistent with what the Supreme Court
21 has said or at least assumed in the past, that if
22 something communicates, it is speech unless it is in
23 some very narrow special category like fraud,
24 defamation, and a few others, incitement.
25 Okay. Well, all right, so data privacy law

1 might have to undergo scrutiny or probably will have
2 to undergo scrutiny. What level of scrutiny is going
3 to apply? This question is much harder to answer. So
4 a case called Dun & Bradstreet vs. Greenmoss Builders
5 involved a credit report, a credit report that was
6 wrong importantly. And in a defamation action, the
7 Supreme Court decided that only intermediate scrutiny,
8 you know, a lower level of protection applies in this
9 defamation case because credit reports that are given
10 to just a couple potential lenders are matters of
11 purely private concern.
12 So you can see a line with this case
13 developing, emerging, that separates speech of public
14 concern or general concern from speech of purely
15 private concern. Dun & Bradstreet could have been
16 limited to just defamation cases, but it has not been
17 limited to that. So the Supreme Court itself has
18 cited to Dun & Bradstreet in cases that have nothing
19 to do with privacy for the proposition that speech of
20 purely private concern is not nearly as protected.
21 So you might think, okay, well, then privacy
22 laws are going to have to only undergo intermediate
23 scrutiny, but more recently, in Reed vs. Town of
24 Gilbert, the Supreme Court decided that strict
25 scrutiny must apply to any law that, on its face,

1 makes a distinction of any sort based on the content
2 of that communication.
3 And if you think about the California
4 Consumer Privacy Act or many of the regulations that
5 the FTC, in the past at least, has considered or that
6 is included in the GDPR, the linchpin for regulation
7 is personal information and it is defined in certain
8 ways and that is all about the content of the data.
9 So if Reed is applied faithfully, it is not clear that
10 courts will be able to do this. But if we are serious
11 about Reed, then it looks like strict scrutiny would
12 apply. At this point, I do not have a confident
13 prediction about which level of scrutiny would apply.
14 But going back to Sorrell for a minute, in
15 the end, when the First Amendment is applied to some
16 sort of privacy law, it is possible that courts could
17 distinguish cases like Sorrell because even in the
18 opinion itself Justice Kennedy said, well, perhaps the
19 state could have addressed physician confidentiality
20 or privacy through a more coherent policy.
21 Now, some might object to the idea that the
22 Vermont law was incoherent because it was targeting
23 kind of the most obnoxious form of data sale, then
24 maybe the legislature was right to just pinpoint that
25 particular form of data sale and leave all others, you

1 know, untampered. But if we take this seriously, then
2 perhaps something like the California statute is more
3 likely to survive because it is broad, because it is
4 so comprehensive.
5 I have some doubts, though, rather that
6 there are at least a few reasons to think that the
7 Government would have to prepare strong arguments and
8 a good base of evidence in order to defend especially
9 a broad privacy law that prohibits the transmission of
10 data. For one thing, just in the past, even since the
11 1960s, the Supreme Court has listened to cases that
12 involve the clash between privacy and the First
13 Amendment, usually in the content of some sort of
14 magazine publication and has found that the privacy
15 interests are not compelling enough to outweigh the
16 general interest in speech.
17 The other thing, though -- I am going to
18 skip this for a second in the interest of time. The
19 other thing is there has been a series of Supreme
20 Court cases, none of them directly related to privacy,
21 but each of them showing that the Supreme Court is
22 extremely skeptical now of any attempt by the
23 Government to justify what it is doing based on just
24 kind of common sense ideas of harms or risks.
25 So Brown vs. Entertainment Merchants

1 Association, for example, was a case that involved a
2 California ban on the sale of violent video games to
3 minors unless the minors had their parents' consent,
4 and the Supreme Court found that the law was
5 unconstitutional, even though the state brought a
6 mountain of social science evidence with it because
7 the Supreme Court -- rightly in my view, but, you
8 know, obviously reasonable minds can differ -- but the
9 Supreme Court thought that social science evidence was
10 actually quite bad. It was poorly done.
11 So the courts are showing an increasing
12 willingness to even look at the -- probe the quality
13 of the evidence that the Government has and offers in
14 order to justify their restrictions on speech.
15 Let me spend just a minute talking about the
16 data collection side of things. So for a long time --
17 so this guy is using his cell phone, I guess, to
18 record someone. So for a long time, the assumption
19 was data collection is not protected by the First
20 Amendment, even though subsequent publication of that
21 information would be.
22 So in a case called Deitemann vs. Time, the
23 Ninth Circuit decided that Time Magazine -- they snuck
24 a couple journalists into a quack's office, like a guy
25 who just was waving wands and turning knobs and

1 pretending to cure diseases, and they did an exposé on
2 him and the Court found that the actual publication
3 was fully protected by the First Amendment. They
4 could not be sued for public disclosure of private
5 facts, irrelevant tort. But the sneaking in of
6 technology to record -- to surreptitiously record what
7 was happening, the secret photographs, that was
8 completely unprotected the Court said.
9 And the Supreme Court, in a case, Bartnicki
10 vs. Vopper, said something similar. That downstream
11 publications are protected, but actually getting
12 access to information or knowledge is unprotected
13 conduct. It is just conduct; it is not speech. That
14 always seemed weird to me because if you think about
15 the reason to limit data collection, it usually has
16 something to do either with knowledge creation by the
17 person who is collecting the data or with downstream
18 communications that that person intends to have.
19 And so if we think of both knowledge and
20 communicating as being core to the First Amendment's
21 goals then why should limitations on collecting
22 information in the first place get a free pass and not
23 get any scrutiny at all?
24 Well, sure enough, in the last couple years
25 -- this is a really recent development, but there have

1 been right-to-record cases that are starting to chip
2 away at this distinction. The first set of cases have
3 to do with recording the police in public. And, now,
4 every circuit that has heard these types of cases has
5 decided that there is a First Amendment right to
6 record police. The Seventh Circuit has gone further
7 and said there is a right to record any time you are
8 in public.
9 And then, also, there have been successful
10 First Amendment challenges to so-called ag-gag laws
11 that prohibit people from secretly recording at
12 commercial farms. And so that, too, is suggesting
13 that even surreptitious recording, even in private
14 spaces, has been getting increasing First Amendment
15 attention.
16 All right. So I raise all of these legal
17 limits not to discourage the FTC in any way from
18 crafting responsible privacy policy, but rather in a
19 way to applaud you for doing these types of hearings
20 because it is tempting to do something like what I
21 think the FDA had done, regrettably, a few years ago
22 and to just kind of plan to defend your policy later
23 in court. But it will save you a lot of headache and
24 heartache if you have a good evidence base and a good
25 theory of what type of interest and seclusion or

1 confidentiality you are actually trying to preserve in
2 order to come prepared for a First Amendment defense.
3 The other option, of course -- and this has
4 come up already -- is to actually prohibit disfavored
5 uses that really are conduct rather than speech. So
6 that is an option as well and then you do not have to
7 defend against the First Amendment at all.
8 All right, thank you very much.
9 (Applause.)
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

1 FTC EXPERIENCE WITH DATA MARKETS
2 MR. STIVERS: All right. So we have Haidee
3 Schwartz as our next speaker.
4 MS. SCHWARTZ: So I am Haidee Schwartz. I
5 am the Acting Deputy Director of the Bureau of
6 Competition at the FTC. First, a disclaimer, these
7 remarks are my own. They are not those of any
8 particular Commissioner or the Commission as a whole.
9 I also want to say that I am looking at this from the
10 Competition side of the FTC. I believe my colleagues
11 in Consumer Protection are talking on other panels.
12 So this is from the Competition side.
13 When people usually talk about data, they
14 usually talk about the four Vs of big data: Volume,
15 how much data are we talking about; velocity, how much
16 data is coming through and how quickly, and for us
17 that means how much does it have to be updated and
18 what is the flow; variety, what are the different
19 forms of data and are they complements or substitutes;
20 and voracity, how accurate or inaccurate is the data?
21 In the FTC context, when we think about data
22 markets, the four Vs are implicitly part of our
23 considerations. And we look at how big data is being
24 used. Is it a product, is it an input, is it a tool?
25 Often, it is two or three of these things. And, of

1 course, we look at whether the data is unique or
2 broadly available. This is particularly important
3 because it affects entry and expansion options of
4 other firms in the market.
5 So how do these cases often look to us and
6 how do they come to us? In the instance of two
7 database companies merging, they often sell data
8 products. And two of the older examples that I have
9 are these type of cases where it is two merging
10 databases. If we go back to 2001, the FTC challenged
11 the consummated merger of Heart Trust and First
12 DataBank. That involved the merger of two competing
13 providers of integral drug data files.
14 Then if you go forward to 2010, the FTC
15 challenged Dun & Bradstreet's acquisition of QED,
16 which was a division of Scholastic that involved K
17 through 12 educational marketing data, such as
18 contact, demographic, and other key information
19 related to teachers, administrators of schools, and
20 school districts. So if you look back, you know, we
21 have a long history of where the database is the
22 product and we are challenge those mergers.
23 In some of the more recent cases I will
24 discuss in this presentation, data was a key input.
25 It wasn't the actual product itself, but it was

1 integral and essential to the product. In many cases
2 we will look at, data is also being used as a tool and
3 it can be a tool and a product, a tool and a key
4 input. And in cases involving data markets, we will
5 look at how the data is being used and whether it is a
6 key differentiator as well as other key dynamics.
7 In these data cases, entry conditions are
8 often critical. What other firms, if any, could
9 replicate the competition lost in their relevant
10 market discussing how data may facilitate or create
11 impediments to that entry.
12 As I have alluded to, the FTC has a long
13 history of cases involving data markets. The history
14 goes back to at least 1996 when the FTC filed
15 administrative complaints again ADP's 1995 acquisition
16 of AutoInfo's assets, charging that the acquisition
17 would raise prices and reduce the quality of service
18 and innovation to the automobile salvage yard
19 information management industry. So these are key
20 tools that the automobile salvage yard used and as
21 well as insurers used. The parties each maintained
22 interchanges which were essentially databases of
23 numbering systems for autoparts and parts assembled
24 that insurers and salvage yard use as sort of an index
25 to determine interchangeability of parts.

1 The parties also had significant software
2 assets, an electronic communication system that
3 allowed auto salvage yards to actually buy the parts
4 and see automatically and quickly, sort of through a
5 central database, what the inventory was at the other
6 yards that subscribed.
7 In the end, the case settled with the
8 divestiture of the former AutoInfo's assets as an
9 ongoing business, which included granting the acquirer
10 an unrestricted license to the interchange which, by
11 that time, had become sort of the default industry
12 standard for a cross-numbering index for parts.
13 Moving on to 2014, CoreLogic and DataQuick,
14 data as a product. This was a merger the FTC
15 challenged in March. In March 2014, CoreLogic agreed
16 to settle FTC charges that its acquisition of
17 DataQuick would likely substantially lessen
18 competition in the market for national assessor and
19 recorder bulk data.
20 So what is national assessor and recorder
21 bulk data? It is current and historical data on
22 properties pulled from local public records, like
23 deeds, mortgages, et cetera, that is aggregated and
24 standardized in bulk format that includes information
25 about ownership value and other characteristics of

1 properties. So it is public information, but it is
2 not standardized, it is not easy to collect, and you
3 need both historical and going forward. Customers of
4 this data, so customers of the companies, use the data
5 in various propriety programs for risk and fraud
6 management tools, valuation models, and a lot of other
7 uses.
8 The complaint alleged that the merger would
9 eliminate one of the three providers of national
10 assessor and recorder bulk data, increasing the risk
11 of coordination between the remaining two firms and
12 the risk that CoreLogic could unilaterally raise
13 prices.
14 In terms of market structure, there were
15 regional assessor and recorder bulk data firms, but
16 the Commission looked at that and saw that they could
17 not combine or reposition to actually compete in the
18 in the national assessor and recorder bulk data
19 market. They would have gaps, they would not be
20 standardized, and there were other issues there.
21 At the time of the merger, CoreLogic
22 licensed its current and go-forward data to DataQuick,
23 which DataQuick was permitted to relicense in bulk.
24 So in other word, DataQuick was actually kind of
25 dependent on CoreLogic for the data. DataQuick used

1 the license data, along with its own historical data,
2 to compete head-to-head with CoreLogic.
3 Importantly, DataQuick was unique in its
4 ability to credibly threaten to enter because it
5 already had historical data. It had acquired a
6 company years before CoreLogic was willing to license
7 to them. Because it had acquired that historical
8 data, CoreLogic viewed it as a potential entrant and,
9 therefore, it sort of got economies of scale and scope
10 by licensing to DataQuick, and it felt that DataQuick
11 would be in there anyway if it did not because it had
12 the historical data. It could have amassed the sort
13 of ongoing data itself. So it was willing to license
14 years ago to DataQuick after it had acquired an
15 historical database.
16 That said, it was very unlikely that anyone
17 else could enter because the breadth of historical
18 data they would need to be gathered and the ability to
19 continue gathering that data would be prohibitive. So
20 no one else was going to have that unique ability to
21 have the historical data.
22 The remedy that we constructed was designed
23 to allow a company called RealtyTrac to step into the
24 shoes of DataQuick as CoreLogic's license. The order
25 required CoreLogic to irrevocably license to

1 RealtyTrac equivalent data to what DataQuick offered
2 to its customers and update the bulk data for five
3 years. That was then designed -- the five years were
4 designed for RealtyTrac to compete with CoreLogic
5 while developing its own ability to collect national
6 bulk data.
7 As we implemented this, RealtyTrac realized
8 that CoreLogic was not providing the entire data set
9 that DataQuick had access to and raised concerns that
10 led to a Commission investigation. Just recently, in
11 March of 2018, the Commission modified the order after
12 finding that CoreLogic had not provided RealtyTrac
13 with all the required data on a timely basis. The
14 modification adds three years to the original term of
15 the order and specifically spells out the quality,
16 service levels, and data transfer requirements.
17 Takeaways from the CoreLogic/DataQuick
18 merger. Here, the data was the product being sold and
19 the breadth, detail, and the complexity of the data
20 created barriers to entry. This matter highlighted
21 the complexities involved in attempting to remedy a
22 lessening of competition when data is the product.
23 You would think it is a database, it is not that hard
24 to transfer, but, here, the buyer’s due diligence may
25 not -- what we learned is the buyer’s due diligence

1 may not necessarily uncover missing or unnecessary
2 data in a timely fashion, and the Commission had
3 difficulty initially identifying the exact universe of
4 data required to effectively compete and required
5 additional work by the buyer, the monitor, and the
6 Commission to determine what data was missing, how it
7 needed to be delivered, and how it needed to be
8 continuously updated.
9 Verisk/EagleView, data as an input. So
10 here, it was not -- in 2014, the Commission issued an
11 administrative complaint seeking to block Verisk's
12 proposed acquisition of EagleView in the growing
13 market for rooftop aerial services. A Verisk
14 subsidiary competed with EagleView to provide software
15 that when combined with the library of aerial images
16 of rooftops allowed insurance adjustors to effectively
17 and efficiently and safely measure roofs.
18 As you can imagine, the old-fashioned way
19 they used to do it was adjustors would actually get up
20 -- well, used to get up on the roofs and do the
21 measurements. That has issues with both accuracy and
22 also significant safety issues. The measurements, in
23 turn, allowed insurers to estimate the cost of repair
24 or replacement of insured roofs. Verisk also owns the
25 software that customers used to make other

1 measurements to estimate damage claims.
2 The Commission alleged a product market of
3 rooftop aerial measurement products, or RAMPs, for
4 insurance purposes. Interestingly, for insurance
5 purposes is key, in terms of if it was actually a
6 targeted customer market because the product was used
7 both by insurers and by adjusters and contractors, but
8 for insurance -- although the software products, you
9 know, functioned somewhat differently, both required
10 the same input, the aerial images and to carry out the
11 same functions.
12 That said, insurance companies -- the
13 Commission judged that insurance companies had
14 different needs and requirements than other customers,
15 like the contractors. You know, the contractors
16 generally felt that they could switch to manual
17 measurements. Insurers could not. As I noted, the
18 product here is not the data itself, but the data was
19 a key input to the product.
20 In terms of the market structure, the merger
21 of these two were judged to create a virtual monopoly.
22 EagleView was the first to develop software using
23 aerial images, and these are actually particular types
24 of aerial images. It is not just any old aerial
25 image. It had to have certain angles, certain types

1 of -- certain types of views, i.e., treeless,
2 leafless. In another few weeks, this will be a good
3 time of year to have aerial images because you can
4 actually see the roof and the measurements. It is not
5 a particular pretty photo, but it does make a
6 difference in terms of aerial photos.
7 And at the time, EagleView had the first
8 mover advantage, amassing a market share of 90
9 percent. It also had, by far, the largest aerial
10 image library. Verisk was a relatively new entrant,
11 entering just two years before the proposed
12 acquisition. But it quickly amassed, you know, a not
13 insignificant market share, substantially more than
14 any other competitor, and it was offering discounts
15 and direct competition to EagleView. The Commission
16 alleged that if the transaction was consummated,
17 discounts would disappear and prices would rise.
18 An important aspect of Verisk and
19 EagleView's competition is their ability to obtain the
20 aerial images that are up-to-date, so the measurements
21 reflected those of current structures, high quality
22 because it allowed adjustors to identify attributes of
23 the insured property, and it also had to be available
24 on a national scale. National insurers wanted to be
25 able to use the software for all of their insured

1 products and it was not worth it to them to sort of
2 have different providers in different areas of the
3 country.
4 EagleView, as I said, had the most extensive
5 library of aerial images. Importantly, insurers also
6 required the RAMPs integrate seamlessly with claims
7 estimation software, and because Verisk was the
8 leading provider of claims estimation software
9 generally, it was able to overcome and was uniquely
10 positioned to be able to overcome a more limited
11 library of aerial images by capitalizing on its
12 relationship with the insurers and the fact that it
13 had the best software and most sort of commonly-used
14 software.
15 Verisk and EagleView abandoned the
16 transaction after the Commission issued the complaint.
17 So the case was never considered by a court. But what
18 the Commission considered in the complaint provides us
19 with some insights. In this case, while data was not
20 the product defined in the product market, it was an
21 essential input into the product and affected a firm's
22 ability to compete and enter the market. The
23 Commission considered the incentives to increase the
24 quality and volume of data as a loss of innovation.
25 So that was also an issue.

1 Now, I am going to talk a little bit about
2 CCC/Mitchell, which was a challenged merger in 2009
3 that the FTC challenged. Full disclosure, I actually
4 was in private practice at the time and was working on
5 behalf of Mitchell, but I am basing this entirely on
6 public information.
7 So access to data as an entry barrier. It
8 was a key input, not the actual product itself. There
9 were two products at issue. One was Estimatics, which
10 is a database used to generate repair estimates for
11 automobiles, this was not particularly the product
12 used for sort of specialized trucks or other things
13 like that, and total loss valuation systems, which
14 were used to determine when a vehicle was totaled and
15 even more importantly, the value of it.
16 At the time of the merger, the big three,
17 which were CCC, Audatex and Mitchell, in that order,
18 had about 99 percent of the estimatics market. There
19 were two fringe competitors. Most importantly, that
20 we will talk about later is Web-Est. And for TLV, the
21 big three accounted for 90 percent of the market.
22 Mitchell had entered later and had a significantly
23 smaller share.
24 There were two types of customers. Insurers
25 and repair facilities for estimatics and primarily

1 insurers for TLV.
2 Okay, database dynamics. So the primary
3 components of estimatics and TLV were the databases
4 themselves and the software. So how did the firms get
5 the databases? CCC had obtained an exclusive license
6 to the Hearst Business Publishing database called
7 "Motor" in 1998. Audatex and Mitchell each had sort
8 of grown their own system painstakingly over years,
9 and part of the reason why Mitchell was smaller is it
10 had taken them many years to create their own
11 database, and they did so.
12 Web-Est licensed Mitchell's database, but
13 under very restrictive conditions. It was not allowed
14 to sell to any of the top 50 insurers, it could not
15 have a communicating product, which meant that
16 basically it could only sell to independent repair
17 stations, not those that were part of a particular
18 repair network, and it could not integrate with other
19 third-party apps, you know, vendors, things that other
20 insurers and other service stations used.
21 So the proposed fix, CCC offered to do two
22 thing in terms of making a database available. One,
23 it offered to relinquish its exclusive rights to the
24 Hearst Motor database. That meant that any new
25 entrant could license that database. And it was fully

1 updated and would continue to be fully updated
2 because Hearst kept that database updated and it was
3 licensed.
4 And Mitchell would remove restrictions on
5 Web-Est and continue that database license. So,
6 therefore, there would be both Web-Est with the
7 Mitchell database and CCC offering to sort of
8 relinquish its exclusive, anyone else could have
9 access to the Hearst database. Audatex would continue
10 with its proprietary-owned database.
11 The judge found the availability of
12 databases would reduce the most critical barrier to
13 entry, but she still found that there was significant
14 other barriers. One, customers were sticky,
15 particularly the insurance customers that were
16 critical to success, and you needed to establish a
17 track record and have a lot of sort of support
18 capabilities. Scale mattered.
19 The judge did note that the Web-Est, which
20 was led by a guy named Eric Seidel had been in the
21 industry for a while, you know, had good experience,
22 had significantly grown his adjustable market share,
23 which had been really independent service stations,
24 but he only had 10 to 15 employees and so the sort of
25 growth curve was going to be too long and to steep.

1 It just would not be sufficient entry in the time
2 required. By comparison to Web-Est, 10 to 15
3 employees, CCC/Mitchell, after they combined, would
4 have had about 2,000 employees.
5 Interestingly, the judge actually decided
6 this case as a coordinated effects case and not as a
7 unilateral effects case. She had found some issues
8 with the FTC's expert's unilateral effects analysis.
9 So it was a PI hearing, not a full trial on the
10 merits, but she decided that the coordinated effects
11 were too likely.
12 Okay. Microsoft and LinkedIn, and this is
13 the last case I am going to discuss before talking
14 about a few takeaways. So Microsoft is obviously
15 strong in operating systems for personal computers and
16 productivity software. LinkedIn is a professional
17 social network that a lot of us probably use. The
18 U.S. investigated, but did not take action. The EC
19 concluded that the merger did not raise competitive
20 concerns related to data, but it did find -- so what
21 it found -- what it looked at -- and these are some of
22 the answers to questions that we often ask -- you
23 know, is the data readily available from other sources
24 or similar data. And, yes, we found that other -- you
25 know, the EC found that other sources existed for that

1 data.
2 They also found that the companies had not
3 particularly provided that data and made it available
4 on the market before. So there was not really going
5 to be a change post-merger. And they had relatively
6 low shares in the market that the EC was concerned
7 about.
8 The EC did require several commitments, and
9 those are just up there. Those primarily had to do
10 with interoperability and ensuring that others could
11 be competitive on the professional social networks.
12 You can see those there. I am not going to read
13 through them. But they did not have to do with
14 particularly the data possessed by the companies.
15 Okay, takeaways, and I am going to try and
16 end early. So takeaways, competition analysis,
17 because I am sure you guys have had a long day and I
18 appreciate you all staying. Current antitrust
19 analysis accounts for how firms compete using data.
20 Data markets and sets are highly differentiated. Each
21 investigation looks very closely at the specific facts
22 of the case. We recognize that data markets are
23 dynamic. Quality and innovation effects may be
24 particularly important. They also may be harder to
25 measure than price effects. How data enables or

1 hinders entry or expansion also may be particularly
2 important.
3 Remedies. In cases that involved data, just
4 like in any other cases, we have a preference for
5 structural remedies. We have seen a number of cases
6 that I have discussed where we look to divest or clone
7 a database versus a license. Sometimes we will allow
8 a license. It depends on the specific facts of the
9 case. There are issues related to how they are going
10 to continue to obtain the data and keep a new data
11 flow that is accurate and is expansive.
12 What we found in our database cases and what
13 we have learned is there is a lot of complexity to how
14 the data is stored, how it is updated, how it is kept
15 and how it is provided to customers. And it seems
16 simple, but there is actually more due diligence that
17 needs to be done not just by buyers of potential
18 assets, but by the Commission and others during that
19 process.
20 There are often IP and copyright issues, and
21 while they are not favored, behavioral conditions may
22 be needed. In some of the cases that I talked about,
23 for example, CoreLogic, there were commitments that we
24 required related to allowing customers to break
25 contracts so that the new firm could have contracts

1 going forward over a certain period of time. So
2 sometimes we have to overcome customer's reluctance
3 and, in some cases, ability to switch. We need to
4 give them the ability to switch, to have the new
5 entrant actually be able to have those customers.
6 There are other types of behavioral
7 conditions, including some support over transition
8 period that we will look at as well. But as noted,
9 structural is always preferred, including in data
10 cases.
11 Thank you, guys.
12 (Applause.)
13 MR. STIVERS: Thanks, Haidee, and thanks all
14 for coming today. We hope we will see many of you
15 back tomorrow morning at 9:00 a.m., and as I think we
16 announced on our website, ultimately there will be a
17 transcript available for these proceedings, as well as
18 the archive webcast. So thanks to all our
19 participants and thanks to all in attendance.
20 (Hearing adjourned.)
21
22
23
24
25

1 CERTIFICATE OF REPORTER
2
3 I, Linda Metcalf, do hereby certify that the
4 foregoing proceedings were digitally recorded by me
5 and reduced to typewriting under my supervision; that
6 I am neither counsel for, related to, nor employed by
7 any of the parties to the action in which these
8 proceedings were transcribed; that I am not a relative
9 or employee of any attorney or counsel employed by the
10 parties hereto, not financially or otherwise
11 interested in the outcome in the action.
12
13
14
15
16 LINDA METCALF, CER
17 Court Reporter
18
19
20
21
22
23
24
25