

**ALGORITHMS: HOW COMPANIES' DECISIONS
ABOUT DATA AND CONTENT IMPACT CONSUMERS**

JOINT HEARING

BEFORE THE

SUBCOMMITTEE ON COMMUNICATIONS AND
TECHNOLOGY

AND THE

SUBCOMMITTEE ON DIGITAL COMMERCE AND
CONSUMER PROTECTION

OF THE

COMMITTEE ON ENERGY AND
COMMERCE

HOUSE OF REPRESENTATIVES

ONE HUNDRED FIFTEENTH CONGRESS

FIRST SESSION

NOVEMBER 29, 2017

Serial No. 115-80



Printed for the use of the Committee on Energy and Commerce
energycommerce.house.gov

U.S. GOVERNMENT PUBLISHING OFFICE

28-578 PDF

WASHINGTON : 2018

COMMITTEE ON ENERGY AND COMMERCE

GREG WALDEN, Oregon

Chairman

JOE BARTON, Texas

Vice Chairman

FRED UPTON, Michigan

JOHN SHIMKUS, Illinois

MICHAEL C. BURGESS, Texas

MARSHA BLACKBURN, Tennessee

STEVE SCALISE, Louisiana

ROBERT E. LATTA, Ohio

CATHY McMORRIS RODGERS, Washington

GREGG HARPER, Mississippi

LEONARD LANCE, New Jersey

BRETT GUTHRIE, Kentucky

PETE OLSON, Texas

DAVID B. MCKINLEY, West Virginia

ADAM KINZINGER, Illinois

H. MORGAN GRIFFITH, Virginia

GUS M. BILIRAKIS, Florida

BILL JOHNSON, Ohio

BILLY LONG, Missouri

LARRY BUCSHON, Indiana

BILL FLORES, Texas

SUSAN W. BROOKS, Indiana

MARKWAYNE MULLIN, Oklahoma

RICHARD HUDSON, North Carolina

CHRIS COLLINS, New York

KEVIN CRAMER, North Dakota

TIM WALBERG, Michigan

MIMI WALTERS, California

RYAN A. COSTELLO, Pennsylvania

EARL L. "BUDDY" CARTER, Georgia

JEFF DUNCAN, South Carolina

FRANK PALLONE, JR., New Jersey

Ranking Member

BOBBY L. RUSH, Illinois

ANNA G. ESHOO, California

ELIOT L. ENGEL, New York

GENE GREEN, Texas

DIANA DeGETTE, Colorado

MICHAEL F. DOYLE, Pennsylvania

JANICE D. SCHAKOWSKY, Illinois

G.K. BUTTERFIELD, North Carolina

DORIS O. MATSUI, California

KATHY CASTOR, Florida

JOHN P. SARBANES, Maryland

JERRY McNERNEY, California

PETER WELCH, Vermont

BEN RAY LUJAN, New Mexico

PAUL TONKO, New York

YVETTE D. CLARKE, New York

DAVID LOEBSACK, Iowa

KURT SCHRADER, Oregon

JOSEPH P. KENNEDY, III, Massachusetts

TONY CARDENAS, California

RAUL RUIZ, California

SCOTT H. PETERS, California

DEBBIE DINGELL, Michigan

SUBCOMMITTEE ON COMMUNICATIONS AND TECHNOLOGY

MARSHA BLACKBURN, Tennessee

Chairman

LEONARD LANCE, New Jersey

Vice Chairman

JOHN SHIMKUS, Illinois

STEVE SCALISE, Louisiana

ROBERT E. LATTA, Ohio

BRETT GUTHRIE, Kentucky

PETE OLSON, Texas

ADAM KINZINGER, Illinois

GUS M. BILIRAKIS, Florida

BILL JOHNSON, Ohio

BILLY LONG, Missouri

BILL FLORES, Texas

SUSAN W. BROOKS, Tennessee

CHRIS COLLINS, New York

KEVIN CRAMER, North Dakota

MIMI WALTERS, California

RYAN A. COSTELLO, Pennsylvania

GREG WALDEN, Oregon (*ex officio*)

MICHAEL F. DOYLE, Pennsylvania

Ranking Member

PETER WELCH, Vermont

YVETTE D. CLARKE, New York

DAVID LOEBSACK, Iowa

RAUL RUIZ, California

DEBBIE DINGELL, Michigan

BOBBY L. RUSH, Illinois

ANNA G. ESHOO, California

ELIOT L. ENGEL, New York

G.K. BUTTERFIELD, North Carolina

DORIS O. MATSUI, California

JERRY McNERNEY, California

FRANK PALLONE, JR., New Jersey (*ex*

officio)

SUBCOMMITTEE ON DIGITAL COMMERCE AND CONSUMER PROTECTION

ROBERT E. LATTA, Ohio

Chairman

ADAM KINZINGER, Illinois

Vice Chairman

FRED UPTON, Michigan

MICHAEL C. BURGESS, Texas

LEONARD LANCE, New Jersey

BRETT GUTHRIE, Kentucky

DAVID B. MCKINLEY, West Virginia

ADAM KINZINGER, Illinois

GUS M. BILIRAKIS, Florida

LARRY BUCSHON, Indiana

MARKWAYNE MULLIN, Oklahoma

MIMI WALTERS, California

RYAN A. COSTELLO, Pennsylvania

GREG WALDEN, Oregon (*ex officio*)

JANICE D. SCHAKOWSKY, Illinois

Ranking Member

BEN RAY LUJAN, New Mexico

YVETTE D. CLARKE, New York

TONY CÁRDENAS, California

DEBBIE DINGELL, Michigan

DORIS O. MATSUI, California

PETER WELCH, Vermont

JOSEPH P. KENNEDY, III, Massachusetts

GENE GREEN, Texas

FRANK PALLONE, JR., New Jersey (*ex*

officio)

C O N T E N T S

	Page
Hon. Robert E. Latta, a Representative in Congress from the State of Ohio, opening statement	2
Prepared statement	3
Hon. Janice D. Schakowsky, a Representative in Congress from the State of Illinois, opening statement	4
Hon. Marsha Blackburn, a Representative in Congress from the State of Tennessee, opening statement	5
Prepared statement	7
Hon. Michael F. Doyle, a Representative in Congress from the Commonwealth of Pennsylvania, opening statement	7
Hon. Greg Walden, a Representative in Congress from the State of Oregon, opening statement	9
Prepared statement	11
Hon. Frank Pallone, Jr., a Representative in Congress from the State of New Jersey, opening statement	12
Prepared statement	13

WITNESSES

Omri Ben-Shahar, Ph.D., Leo Herzel Professor in Law, University of Chicago Law School	15
Prepared statement	18
Answers to submitted questions	154
Kate Klonick, Resident Fellow, Information Society Project, Yale Law School	23
Prepared statement	25
Answers to submitted questions ¹	157
Laura Moy, Deputy Director, Center on Privacy & Technology at Georgetown Law	33
Prepared statement	35
Answers to submitted questions	159
Catherine Tucker, Ph.D., Sloane Distinguished Professor of Management Science, MIT Sloane School of Management	57
Prepared statement	59
Answers to submitted questions	164
Frank Pasquale, Professor of Law, University of Maryland	67
Prepared statement	69
Answers to submitted questions	169
Michael Kearns, Ph.D., Computer and Information Science Professor, Univer- sity of Pennsylvania	93
Prepared statement	95
Answers to submitted questions	183

SUBMITTED MATERIAL

Letter of November 1, 2016, from Hon. Robin L. Kelly, et al., to Mark Zuckerberg, Chairman and Chief Executive Officer, Facebook, submitted by Ms. Clarke	130
Statement of Frank Pasquale, Professor of Law, University of Maryland, before the United States Senate, September 12, 2017, submitted by Ms. Matsui	131

¹Ms. Klonick did not answer submitted questions for the record by the time of printing.

VI

	Page
Tweets of November 22, 2017, Cloudflare, submitted by Mr. Costello	149
Report of May 2016, "Online Privacy and ISPs," The Institute for Information Security & Privacy, ¹ submitted by Mr. Lance	
Letter of November 28, 2017, from Marc Rotenberg, President, and Caitriona Fitzgerald, Policy Director, Electronic Privacy Information Center, to Mr. Latta, et al., submitted by Mr. Lance	150

¹The information has been retained in committee files and also is available at <http://docs.house.gov/meetings/IF/IF17/20171129/106659/HHRG-115-IF17-20171129-SD004-U4.pdf>.

ALGORITHMS: HOW COMPANIES' DECISIONS ABOUT DATA AND CONTENT IMPACT CON- SUMERS

WEDNESDAY, NOVEMBER 29, 2017

HOUSE OF REPRESENTATIVES,
SUBCOMMITTEE ON COMMUNICATIONS AND TECHNOLOGY
JOINT WITH THE
SUBCOMMITTEE ON DIGITAL COMMERCE AND CONSUMER
PROTECTION,
COMMITTEE ON ENERGY AND COMMERCE,
Washington, DC.

The subcommittee met, pursuant to call, at 10:07 a.m., in room 2123, Rayburn House Office Building, Hon. Robert E. Latta (chairman of the Subcommittee on Digital Commerce and Consumer Protection) presiding.

Members present: Representatives Latta, Blackburn, Harper, Lance, Shimkus, Burgess, Guthrie, Olson, Kinzinger, Bilirakis, Johnson, Bucshon, Flores, Brooks, Mullin, Collins, Cramer, Walters, Costello, Walden (ex officio), Doyle, Schakowsky, Eshoo, Engel, Green, Matsui, McNerney, Welch, Clarke, Loeb sack, Ruiz, Dingell, and Pallone (ex officio).

Staff present: Mike Bloomquist, Deputy Staff Director; Samantha Bopp, Staff Assistant; Kelly Collins, Staff Assistant; Robin Colwell, Chief Counsel, Communications and Technology; Sean Farrell, Professional Staff Member, Communications and Technology; Margaret T. Fogarty, Staff Assistant; Melissa Froelich, Chief Counsel, Digital Commerce and Consumer Protection; Adam Fromm, Director of Outreach and Coalitions; Gene Fullano, Detailee, Communications and Technology; Ali Fulling, Legislative Clerk, Oversight and Investigations, Digital Commerce and Consumer Protection; Theresa Gambo, Human Resources and Office Administrator; Elena Hernandez, Press Secretary; Paul Jackson, Professional Staff Member, Digital Commerce and Consumer Protection; Bijan Koohmaraie, Counsel, Digital Commerce and Consumer Protection; Tim Kurth, Senior Professional Staff, Communications and Technology; Lauren McCarty, Counsel, Communications and Technology; Katie McKeogh, Press Assistant; Alex Miller, Video Production Aide and Press Assistant; Mark Ratner, Policy Coordinator; Madeline Vey, Policy Coordinator, Digital Commerce and Consumer Protection; Evan Viau, Legislative Clerk, Communications and Technology; Hamlin Wade, Special Advisor for External Affairs; Everett Winnick, Director of Information Technology; Greg Zerzan, Counsel, Digital Commerce and Consumer Protection;

Michelle Ash, Minority Chief Counsel, Digital Commerce and Consumer Protection; Jeff Carroll, Minority Staff Director; David Goldman, Minority Chief Counsel, Communications and Technology; Lisa Goldman, Minority Counsel; Lori Maarbjerg, Minority FCC Detailee; Dan Miller, Minority Policy Analyst; Caroline Paris-Behr, Minority Policy Analyst; and C.J. Young, Minority Press Secretary.

**OPENING STATEMENT OF HON. ROBERT E. LATTA, A
REPRESENTATIVE IN CONGRESS FROM THE STATE OF OHIO**

Mr. LATTA. Well, good morning. I would like to call our joint subcommittee meeting to order, and the Chair recognizes himself for 5 minutes for an opening statement.

And good morning again. I would like to welcome everyone back from Thanksgiving holiday to our joint subcommittee hearing. I would like to thank our witnesses for being here today. I would venture to guess many people were able to get a jumpstart on their holiday shopping and seeing some of the earlier reports showing that online shopping rose 17 percent from last year, which makes our hearing this morning even more timely.

When Chairman Walden became chairman of the Energy and Commerce Committee, we agreed that keeping our focus on the consumer was a priority for the committee. And everything that the Digital Commerce and Consumer Protection Subcommittee has done, whether it has been exploring new technologies through our Disrupter Series or the bipartisan work that went into the SELF DRIVE Act, our goal has always been to act in the best interest of the consumer, the American people.

Earlier this fall, the Equifax data breach compromised the personal information of over 145 million Americans. This troubling incident raised many questions about credit industry practices with respect to the collection of consumer information. Many Americans, some of whom never heard of Equifax, were confused as to how their sensitive personal information could have been compromised by a company they had never interacted with.

Just last week, Uber announced their systems were hacked, exposing data of over 57 million users. Rather than alert authorities and make the breach known to their users and drivers, Uber kept the hack secret for a year. Disregard of law and disregard of consumers' and drivers' trust all require close scrutiny. The Digital Commerce and Consumer Protection Subcommittee will continue our work to protect consumers and make sure those who disregard the law are held accountable.

As investigations continue, the importance of this hearing cannot be understated. Polls show Americans both feel that technology has had a positive effect on our society but are also skeptical about how their information is used by major technology companies. As policymakers, it is our obligation to ask the tough questions and make sure consumers understand how their information is being used in our digitally driven economy.

That is why we explore today how personal information about consumers is collected online and, importantly, how companies use that information to make decisions about the content consumers see. Right now, there are more than 224 million smart phone users in America, and U.S. consumers spend about 5 hours a day on

their mobile devices. As we continue to see the number of connected devices increase and our digital economy expand, Americans are only going to spend more and more time online browsing the web, shopping, or checking social media, with more information about them being collected.

Although there are legitimate reasons and benefits of the collection and use of information online, we want to ensure that Americans understand how their information is being used. Specifically, how do companies use algorithms to make decisions and deliver content to consumers? What information goes into these complex algorithms, and how do they control the information that comes out? How important are human decisions in creating the algorithms and interpreting the results? Are the results of the researches we conduct online objective, or are companies controlling the information we get?

These are all fair, legitimate questions that we intend to explore. It is our job to make sure consumers have the information they need to make informed decisions, especially when it comes to the flow of their personal information online. With that said, it is also important to understand how effective privacy policy disclosures are. Although some scholars believe such disclosures empower the consumers, others contend they are only there for the lawyers and are impossible to read. For that reason, we must consider whether there are more effective ways to empower the consumer.

I would like to thank Chairman Blackburn for her commitment to these issues, and I look forward to exploring these complex but important issues with all stakeholders. Again, I want to thank our witnesses for being here today, and at this time I would like to recognize the gentlelady from Illinois, the ranking member of the subcommittee, for 5 minutes.

[The prepared statement of Mr. Latta follows:]

PREPARED STATEMENT OF HON. ROBERT E. LATTA

Good morning, I'd like to welcome everyone back from the Thanksgiving holiday to our joint subcommittee hearing. I'd like to thank our witnesses for being here today. I would venture to guess many people were able to get a jump start on their holiday shopping. Early reports show online shopping revenues rose over 17 percent from last year, which makes our hearing this morning so timely.

When Chairman Walden became chair of the Energy and Commerce Committee, we agreed that keeping our focus on the consumer was a priority for the committee. In everything the Digital Commerce and Consumer Protection subcommittee has done—whether it has been exploring new technologies through our Disrupter Series or the bipartisan work that went into the SELF DRIVE Act—our goal has always been to act in the best interest of the consumer and the American people.

Earlier this fall, the Equifax data breach compromised the personal information of over 145 million Americans. This troubling incident raised many questions about credit industry practices with respect to the collection of consumer information. Many Americans—some of who had never heard of Equifax—were confused as to how their sensitive personal information could have been compromised by a company they had never interacted with.

Just last week, Uber announced their systems were hacked exposing data on over 57 million users. Rather than alert authorities and make the breach known to their users and drivers—Uber kept the hack secret for a year. Disregard of the law and disregard of consumers and drivers trust all require close scrutiny. The Digital Commerce and Consumer Protection subcommittee will continue our work to protect consumers and make sure those who disregard the law are held accountable.

As investigations continue, the importance of this hearing cannot be understated. Polls shows Americans both feel that technology has had a positive effect on our society, but are also skeptical about how their personal information is used by major

technology companies. As policymakers, it is our obligation to ask the tough questions and make sure consumers understand how their information is being used in our digitally driven economy.

That is why we will explore today how personal information about consumers is collected online and—importantly—how companies use that information to make decisions about the content consumers see.

Right now, there are more than 224 million smartphone users in America and U.S. consumers spend about 5 hours a day on their mobile devices. As we continue to see the number of connected devices increase and our digital economy expand, Americans are only going to spend more and more time online—browsing the web, shopping, or checking social media—with more information about them being collected.

Although there are legitimate reasons and benefits to the collection and use of information online, we want to ensure that Americans understand how their information is being used.

Specifically, how do companies use algorithms to make decisions and deliver content to consumers? What information goes into these complex algorithms and how do they control the information that comes out? How important are human decisions in creating the algorithms and interpreting their results? Are the results of the searches we conduct online objective or are companies controlling the information we get? These are all fair, legitimate questions that we intend to explore.

It is our job to make sure consumers have the information they need to make informed decisions—especially when it comes to the flow of their personal information online. With that said, it is also important to understand how effective privacy policy disclosures are. Although some scholars believe such disclosures empower the consumer, others contend that they are only there for the lawyers and are impossible to read. For that reason, we must consider whether there are more effective ways to empower the consumer.

I would like to thank Chairman Blackburn for her commitment to these issues, and I look forward to exploring these complex, but important issues with all stakeholders.

Thank you again to our witnesses for being here today.

OPENING STATEMENT OF HON. JANICE D. SCHAKOWSKY, A REPRESENTATIVE IN CONGRESS FROM THE STATE OF ILLINOIS

Ms. SCHAKOWSKY. Thank you, Mr. Chairman. We like to think of the internet as an open marketplace and forum for the exchange of ideas. In reality, the information that consumers see is determined in part by tech companies. Today, algorithms determine what appears in web ads, search results, and your customized news feed. Some of the content you are presented may be based on personal information such as your gender, race, and location. It may also depend on how much companies have paid to get that content in front of you.

The internet and social media have changed how Americans consume news, information, and advertising. According to an August 2017 survey by the Pew Research Center, two-thirds of Americans get at least some of their news through social media. Consumers rely on a handful of popular platforms, making the algorithms of those platforms tremendously powerful.

On a sinister level, organizations and even nation-states can exploit algorithms to spread disinformation, as we saw with Russian interference in the 2016 elections. In addition, platforms profit by selling ads targeted to specific groups based on their demographics and inferences made through their engagement with content on the platform. This may have some benefit: Consumers see ads that they are actually interested in. But the line between tailoring advertising and facilitating discrimination can get murky.

As we grapple with algorithms on the internet, the Federal Communications Commission is considering big changes that would allow corporations to further shape what content consumers access. On December 14th, the FCC will vote on whether to undo the Open Internet Order, which protects net neutrality. If that proposal is adopted, internet service providers will be able to control consumers' access to content. They can make a website load faster or slower depending on whether the content provider pays for the better speed, or an ISP can block content altogether.

Destroying that neutrality would change the internet as we know it, and how does a small business compete online if it now has to pay every ISP in the country for its website to load as fast as big corporation competitors? What happens to the exchange of ideas when access to some content is restricted? This is a disturbing amount of power that the FCC might cede to for-profit broadband providers.

We already have examples of what broadband providers do when empowered to block content. Verizon blocked text messages from reproductive rights group NARAL, calling them, quote, controversial, unquote. AT&T limited use of FaceTime to incentivize its customers to purchase more expensive data plans. TELUS, another telecom company, blocked the website of a union with which it had a labor dispute. No wonder millions of internet users have filed comments in support of maintaining the Open Internet Order. Just since last Monday, my office has received about 500 calls from net neutrality supporters.

Americans are watching the FCC's next move. The FCC under Chairman Pai is also encouraging consolidation and media ownership. It has bent over backward to clear the way for Sinclair Broadcast Group's acquisition of Tribune Media. Congress established a 39 percent cap on the national audience one broadcaster can cover, but Chairman Pai moved to reinstate the outdated UHF discount so that Sinclair can potentially cover 70 percent of the national audience. This media consolidation is a threat to local journalism, especially as Sinclair forces its stations to run nationally produced, quote, must-air, unquote, content.

Big corporations are being given more and more influence over the information that Americans receive, from news feeds to websites, from smart phone to TVs. Congress and Federal watchdogs like the FCC have a responsibility to push back on corporate power when it threatens fair competition and free expression. I look forward to our witnesses' insights on how we fulfill that responsibility, and I yield back. Thank you.

Mr. LATTA. Thank you. The gentlelady yields back, and at this time the Chair recognizes the gentlelady from Tennessee, the chairman of the Communications and Technology Subcommittee, for 5 minutes.

OPENING STATEMENT OF HON. MARSHA BLACKBURN, A REPRESENTATIVE IN CONGRESS FROM THE STATE OF TENNESSEE

Mrs. BLACKBURN. Good morning, and welcome to all of our witnesses. I want to thank my colleague Mr. Latta for working closely

with me and our committee to put together this stellar panel so that we can talk about all things virtual.

Although we often refer to the world on the other side of the screen as the virtual world, we are seeing that, when things go wrong, the real-world impacts on our privacy, finances, knowledge base, and even freedom of expression are anything but virtual. They are very, very real. As so many of these issues overlap between our two subcommittees, I am pleased we are able to kick off our exploration of these issues as a team.

On a number of fronts, we are seeing the pressure turned up on the tech companies that often serve as the new town squares for public discourse as governments and users are demanding that certain speech be shut down. Some of the responses have perhaps been a disappointment from the perspective of free speech. Companies that began as start-ups in Silicon Valley garages have fundamentally changed the way we communicate with one another about everything from the song we want to hear, to what stock to buy, to what is the best way to change our healthcare delivery system.

These multinational corporations now respond to pressures that do not necessarily align with American values, so we need to examine how and why content is being blocked, filtered, or prioritized. This may all sound faintly similar to another topic, net neutrality. Exercise caution here, as it is important to note the FCC's current rules only apply to ISPs, not social media or search platforms.

In some very concrete ways, the open internet is being threatened by certain content management practices. These 2-year-old FCC rules have not and cannot address these threats, so it is disheartening to see Title II regulatory advocates happily conflating the two to divert attention from who is actually blocking content. The current FCC proposal to return internet regulation back to the bipartisan light-touch norm also reminds us that we are simply shifting authority back to the FTC to handle privacy matters.

The previous head of the FCC swiped jurisdiction from the FTC, a 100-plus-year-old institution established by a Democratic President to act against trusts. As discussed at our previous hearings on the limits of the FCC, its authority can only touch one part of the internet ecosystem, and thus it ignores edge provider services that collect arguably more data than ISPs.

As you may have heard, in order for consumers to be able to protect their virtual you, I introduced a bill that would create a level and fair privacy playing field by bringing all entities that collect and sell personal data of individuals under the same unified rules. Given the witnesses' testimony today, let me also plug another bipartisan initiative we have addressed: data security. Given the implications and risk associated with transferring all of this data, it is imperative that we address data security. It is a timely issue.

I look forward to working with my friends across the aisle on this, data security, and on privacy, the BROWSER Act, and all of these topics so that we can settle our differences right here with legislative authority in these hearing rooms rather than relinquishing that authority to regulators in power. I thank the chairman for his collaboration and work on this issue, and I yield back the balance of my time.

[The prepared statement of Mrs. Blackburn follows:]

PREPARED STATEMENT OF HON. MARSHA BLACKBURN

Good afternoon, and welcome to our witnesses. Let me also thank my colleague Mr. Latta for working closely with me to put together this all-star panel to discuss all things virtual. Although we often refer to the world on the other side of our screens as the virtual world, we are seeing that when things go wrong, the real world impacts on our privacy, finances, knowledge base, and even freedom of expression are anything but virtual. As so many of these issues overlap between our two subcommittees, I am pleased that we are able to kick off our exploration of them as a team.

On a number of fronts, we are seeing the pressure turned up on the tech companies that often serve as the new town squares for our public discourse. As governments and users are demanding that certain speech be shut down, some of the responses have perhaps been a disappointment from the perspective of free speech. Companies that began as start-ups in Silicon Valley garages have fundamentally changed the way we communicate with each other about everything from what song we want to hear, to what stock we want to buy or sell, to what is the best way to change our health care system. These multinational corporations now respond to pressures that do not necessarily line up with American values, so we need to examine how and why content is being blocked, filtered, or prioritized.

This may all sound faintly similar to another hot topic—net neutrality. Exercise caution here as it is important that we note: the FCC's current rules only apply to ISPs, not social media or search platforms. In some very concrete ways, the open internet is being threatened by certain content management practices. These 2-year-old FCC rules have not and cannot address these threats, so it is disheartening to see Title 2 regulatory advocates happily conflating the two to divert attention from who is actually blocking content.

The current FCC proposal to return internet regulation back to the bipartisan light-touch norm also reminds us that we are simply shifting authority back to the FTC to handle privacy matters. The previous head of the FCC swiped jurisdiction from the FTC, a 100-plus-year-old institution established by a Democratic president to act against trusts. As discussed at our previous hearings on the limits of the FCC, its authority can only touch one part of the internet, ecosystem and thus it ignores edge provider services that collect arguably more data than ISPs. As you may have heard, I introduced a bill that would create a level and fair privacy playing field by bringing all entities that collect and sell the personal data of individuals under the same rules.

Given the witnesses testimony today, let me also plug another bipartisan initiative I have worked on in the past—data security. Given the implications and risks associated with transferring all of this data, it feels rather timely. I look forward to working with my friends across the aisle on this and all of these topics so we settle differences in this hearing room as opposed to relinquishing our authority to regulators in power.

Mr. Latta. Thank you. The gentlelady yields back. And, before I recognize our next Member, I just want to mention to our witnesses we have another subcommittee that is going on right now, so you will have Members coming in and out of subcommittee today. And at this time, the Chair recognizes the gentleman from Pennsylvania, the ranking member on C&T, for 5 minutes.

OPENING STATEMENT OF HON. MICHAEL F. DOYLE, A REPRESENTATIVE IN CONGRESS FROM THE COMMONWEALTH OF PENNSYLVANIA

Mr. Doyle. Thank you, Mr. Chairman, for holding this joint hearing, and thank you to the witnesses who have come before us today.

Machine learning and artificial intelligence are powerful tools that are reshaping our country and our economy. In places like my hometown of Pittsburgh, our leadership in artificial intelligence is leading to new technologies and new advances that have the poten-

tial for revolutionary changes. I hope this committee can continue to investigate and understand this important technology and the impacts that it will have.

That being said, troubling recent events such as the hack of Equifax continue to show light on the dark world of data brokers and data mining. Credit rating agencies play a central role in many Americans' lives whether you are buying a home, a car, or even a new phone. Your ability to demonstrate good credit in the eyes of these institutions is tantamount to being allowed to make a purchase or being told that you do not pass Go. Americans have little recourse, and our Government provides little oversight of these institutions and their practices. They are increasingly using big data and machine learning to make judgments about individuals and their ability to access and use credit.

Data breaches at these companies pose grave threats to nearly every American, and I think this warrants further investigation. However, today I am deeply concerned that this hearing is happening in the shadow of the FCC's efforts to end network neutrality and this Congress' own decision to use the Congressional Review Act on the FCC's broadband privacy rules. These policies are and were robust protections for consumers that are at the heart of our discussions here today.

In addition, Ms. Moy's testimony refers in numerous places to the CRA against rules requiring mandatory arbitration by financial institutions. The majority does not seem content to merely strip Americans of their legal and regulatory protections. They are going even further now and working to deny them their access to the courts, as well. The majority seems willing only to give lip service to these real consumer protections that they have already cast aside.

The FCC's current efforts to repeal the Open Internet Order and end network neutrality are a perfect case in point. The need for net neutrality was borne out of a long history of anti-consumer and anti-competitive behavior that limited consumers' access to content and information, new technologies, and competitive choices. ISPs have blocked consumer access to services that compete with their own services, new services, and transformative services more times than I can count. The FCC's privacy rules themselves were a reaction to bad behavior by the ISPs.

For years, ISPs have taken actions to track user behavior online using deep packet inspection, undeletable supercookies, and even force consumers to pay them on top of the sky-high fees they already charge to retain their privacy. Consumers were protected from these abusive practices until Congress and President Trump recklessly acted to nullify these rules.

I cannot reiterate to my colleagues enough that when you own the pipe to the home, you own access to the consumer, as ISPs have demonstrated so many times. Repealing these rules will have grave consequences on consumers and the vibrance of the online ecosystem. I continue to urge Chairman Pai to end his quixotic misadventure, and with that being said, I will yield the remainder of my time to Mr. McNerney.

Mr. MCNERNEY. I thank the ranking member. While I am glad we are holding today's hearing about protecting online consumers,

I am disappointed that the Republicans on this committee and at the Federal Communications Commission are doing just the opposite. Earlier this year, Republicans passed the privacy CRA, eliminating broadband privacy protections for consumers' personal information.

In response, I introduced the MY DATA Act. This legislation would give the Federal Trade Commission rulemaking and enforcement authority so that consumers can have strong privacy and data security protections across the internet. Not a single Republican agreed to cosponsor this bill. In addition, this December, the FCC is expected to adopt Chairman Pai's proposal to dismantle net neutrality.

Thousands of constituents have reached out to my office this year to express concerns about eliminating broadband privacy and net neutrality protections. I urge my Republican colleagues to take actions to actually protect consumers instead of talking about protecting consumers while exposing consumers to online mischief. I yield back.

Mr. LATTI. Thank you very much. The gentleman yields back. And at this time, the Chair now recognizes the gentleman from Oregon, the chairman of the full committee, for 5 minutes.

OPENING STATEMENT OF HON. GREG WALDEN, A REPRESENTATIVE IN CONGRESS FROM THE STATE OF OREGON

Mr. WALDEN. Thank you, Mr. Latta, and good morning, everyone. Thanks for being here, especially thanks to our witnesses.

And today we begin a critical discussion about the evolution of consumers' online environment. We will dive into many of the important questions surrounding the future of data access and content management in a marketplace driven by algorithms. Just in the past decade, the internet economy has grown, thrived, and evolved, as you all know, substantially. It is amazing what is happening there.

The smart phones we carry with us everywhere, the tablets we log on to, the smart home devices in our kitchens, all represent a transformational shift in how Americans gather information, receive their news and content, and how they connect with friends and with family. These services are convenient, efficient, and provide valuable and tangible benefits to American consumers.

The companies behind the services have created thousands and thousands of jobs and brought the U.S. into the forefront of technology and innovation. In exchange for using certain websites or platforms, consumers are willing to share personal details about themselves—names, locations, interests, and more. The context of the relationship drives that exchange.

Now, depending on the service, tech companies and online platforms make their money because they know who you are, where you are, what you like, what photos and videos you take and watch, and what news you read. The depth and power of data will be supercharged with the proliferation of connected and embedded devices in the Internet of Things. Billions of IoT devices will surely be deployed, linking machines to other machines and transmitting massive amounts of data and information to connect Americans to

even more services, conveniences, and benefits from all around the globe.

So what is behind these services and activities? Algorithms and data. Algorithms are a sequence of instructions to solve a problem or complete a task. These instructions help devices and apps predict user preferences as well as provide the content and advertising you see in your social media feed. Data serve as inputs or signals to those algorithms. Well-intentioned algorithms can lead to unanticipated consequences. For example, algorithmic bots are being profusely designed to steal or to cheat in online gambling and ticket sales.

Humans remain a critical part of the creation and monitoring of these systems. In recent months, reports of data breaches and algorithms gone awry have demonstrated the potentially negative influences of digital technology on Americans' lives. This committee has done extensive work on issues surrounding consumer protection and data breaches. We brought in the former CEO of Equifax for a hearing, and we continue to push for answers on behalf of American consumers.

At the same time, there have been some high-profile instances of major social media platforms blocking content for questionable reasons using opaque processes. As a result of all of this, consumers are concerned about whether they can trust online firms with the integrity of news and information they disseminate, the welfare of its users, and on a much larger scale the preservation of our own democratic institutions. All these are part of the big public discussion going on right now.

As we all know net neutrality is the issue of the moment, but regardless of where you stand on that policy, the recent attacks on Chairman Pai and particularly his children are completely unacceptable and have no place in this debate. Period. I condemn it in the strongest terms, and I call on the entire tech community and my colleagues on both sides of the aisle to condemn it, as well.

In light of the current controversy surrounding net neutrality rules for ISPs, it is important to examine how content is actually being blocked or promoted or throttled every day on the internet and not by the ISPs. Net neutrality rules do not address the threats to the open internet that we will discuss today.

Now, the goal for today's hearing is to help provide all Americans with a better understanding of how their data flows online, how online platforms and online media sources determine what they see or don't see, and the extent of and methods by which their information is collected and used by online firms. Americans should be able to feel confident that their well-being, freedom of expression, and access to the content of their choice are not being wholly sacrificed for profit.

Americans should have vibrant, competitive markets both offline and online where consumers know their rights and options and have the freedom to choose what is best for their circumstances. It is undeniable the internet has created millions of new jobs, tremendous opportunities, access in ways unimaginable just a few years ago, but it has also created these new risks and challenges.

So, in the name of convenience, is there a potential for online firms to undermine America's privacy and security in a way that

they don't expect or know about? Are the current policies regarding the collection and use of personal data working? Are consumers harmed by this hyperpersonalization? And finally, are firms' content management practices constraining America's ability to speak and to listen freely on an open internet?

Consumers should remain as safe from unfair, deceptive, and malicious practices by online firms and their algorithms on the internet as they do in the real world. And we are here today to dig into these tough questions, and we appreciate your advice and counsel from our witnesses today. And with that, Mr. Chair, I yield back.

[The prepared statement of Mr. Walden follows:]

PREPARED STATEMENT OF HON. GREG WALDEN

Good morning. Today we begin a critical discussion about the evolution of consumers' online environment. We will dive into many important questions surrounding the future of data access and content management in a marketplace driven by algorithms.

Just in the past decade, the internet economy has grown, thrived, and evolved substantially. The smartphones we carry with us everywhere, the tablets we log on to, and the smart home devices in our kitchens all represent a transformational shift in how Americans gather information, receive news and content, and connect with friends and family.

These services are convenient, efficient, and provide value and tangible benefits to American consumers. The companies behind the services have created jobs, and brought the U.S. into the forefront of technological innovation.

In exchange for using certain websites or platforms, consumers are willing to share personal details about themselves—names, locations, interests, and more. The context of the relationship drives that exchange.

Depending on the service, tech companies and online platforms make their money because they know who you are, where you are, what you like, what photos and videos you take and watch, and what news you read.

The depth and power of data will be supercharged with the proliferation of connected and embedded devices in the Internet of Things.

Billions of IoT devices will surely be deployed, linking machines to other machines, and transmitting massive amounts of data and information to connect Americans to even more services, conveniences and benefits from all around the globe.

What's behind these services and activities? Algorithms and data.

Algorithms are a sequence of instructions to solve a problem or complete a task. These instructions help devices and apps predict user preferences as well as provide the content and advertising you see in your social media feed. Data serve as inputs or signals to the algorithms.

Well-intentioned algorithms can lead to unanticipated consequences. For example, algorithmic bots are being purposefully designed to steal or to cheat in online gambling and tickets sales. Humans remain a critical part of the creation and monitoring of these systems.

In recent months, reports of data breaches and algorithms gone awry have demonstrated the potentially negative influences of digital technology on Americans' lives.

This committee has done extensive work on issues surrounding consumer protection and data breaches—we brought in the former CEO of Equifax for a hearing—and we continue to push for answers on behalf of consumers.

At the same time, there have been some high-profile instances of major social media platforms blocking content for questionable reasons, using opaque processes.

As a result of all this, consumers are concerned whether they can trust online firms with the integrity of the news and information they disseminate, the welfare of its users, and, on a much larger scale, the preservation of our democratic institutions.

As we all know, net neutrality is the issue of the moment, but regardless of your position on the policy, the recent attacks on Chairman Pai and particularly his children, are completely unacceptable and have no place in this debate. I condemn it in the strongest terms and I call on the entire tech community and my colleagues on both sides of the aisle to condemn it as well.

In light of the current controversy surrounding net neutrality rules for ISPs, it's important to examine how content is actually being blocked and throttled every day on the internet—and not by the ISPs.

While I will continue to pursue legislation on net neutrality rules, the fact is, they do not and cannot address the threats to the open internet that we will discuss today.

The goal for today's hearing is to help provide all Americans with a better understanding of how their data flows online, how online platforms and online media sources determine what they see or don't see, and the extent of and methods by which their information is collected and used by online firms.

Americans should be able to feel confident that their well-being, freedom of expression, and access to the content of their choice are not being wholly sacrificed for profit.

Americans should have vibrant, competitive markets both offline and online, where consumers know their rights and options, and have the freedom to choose what is best for their circumstances.

It is undeniable the internet has created new jobs, tremendous opportunity, and access in ways unimaginable just a few years ago. But it has also created new risks and challenges.

In the name of convenience, is there the potential for online firms to undermine Americans' privacy and security in a way that they don't expect?

Are the current policies regarding the collection and use of personal data working? Are consumers harmed by this hyper-personalization?

And finally, how are firms' content management practices constraining Americans' ability to speak and to listen freely on an open internet?

Mr. LATTI. Thank you very much. The gentleman yields back, and at this time the Chair recognizes the gentleman from New Jersey, the ranking member of the full committee, for 5 minutes.

OPENING STATEMENT OF HON. FRANK PALLONE, JR., A REPRESENTATIVE IN CONGRESS FROM THE STATE OF NEW JERSEY

Mr. PALLONE. Thank you, Mr. Chairman. The internet is home to some of the most important conversations taking place today. As internet companies find ways for Americans to communicate, our democracy should be stronger than ever, but as you all know something else is going on. Our national dialogue is being curated by companies policing content, and the number of websites handling this traffic has consolidated to just a few key players.

The aim of internet platforms is monetizing web traffic, not public policy. Algorithms created for the purpose of increasing ad clicks is what ends up shaping what we see online, and too often this content is not an accurate reflection of the real world. Structural flaws built into the algorithms used to sort online content may result in racial and other bias in our news feeds.

As diverse voices are squeezed out, bias increases even further, and this is simply not acceptable, and I look forward to hearing more today about what we can do about it. Unfortunately, forces are at work here in Washington that make this problem worse. At every turn, we see efforts to give more power to gatekeepers, either by eviscerating net neutrality and privacy or by picking favorite voices for preferred regulatory treatment.

Even now, as we hold a hearing to talk about mitigating bias on the internet, FCC Chairman Pai is planning to introduce more bias into the system. The net neutrality rules that he plans to destroy are the protections that ensure that we the people can decide for ourselves what we do and say online, and Chairman Pai's plan will fundamentally change the free and open internet as we know it.

Independent voices, those outside the mainstream, may be most at risk simply because they don't have an affiliation with the companies that run the internet.

Unfortunately, broadband companies have more than just financial reasons to obstruct access to independent content, it can also be political. Under Chairman Pai's plan, nothing stops those in power from pushing broadband companies to censor dissenting voices or unpopular opinions or to promote views that they support. We are seeing more and more often how this administration is using its political might to pressure even large companies.

And this is not a partisan point or even a political one. Jeopardizing the national dialogue should concern all of us. The dialogue that happens online is critical for our democracy. Chairman Pai's move comes after this Congress acted earlier this year to wipe out privacy and data security online. Under President Obama, the FCC adopted fair rules to protect the little guy: ask before collecting information, don't share it without consent, and take reasonable measures to safeguard it. But that was too much for congressional Republicans who voted to take away these protections and hand over consumers' data to big business.

Sadly, there is still more to come. Over this past year, the FCC has taken every step possible to ensure that Sinclair broadcasting, already the largest owner of broadcast stations in the country, becomes even bigger. And these steps by the FCC fly in the face of laws Congress put in place to protect local voices. We understand that diverse perspectives are critical for our communities and strengthen our democracy. Instead, the FCC is doing everything it can to allow one company to control what people hear no matter where they are in the country and that is simply not what we intended.

So I look forward to discussing ways to eliminate bias in our communication systems. We need to figure out how to wrest power over information from corporations and return it back to the people. And I yield the remainder of my time to the gentlewoman from New York, Ms. Clarke.

[The prepared statement of Mr. Pallone follows:]

PREPARED STATEMENT OF HON. FRANK PALLONE, JR.

The internet is home to some of the most important conversations taking place today. As internet companies find ways for Americans to communicate, our democracy should be stronger than ever. But as we all know, something else is going on. Our national dialogue is being curated by companies policing content, and the number of websites handling this traffic has consolidated to just a few key players.

The aim of internet platforms is monetizing web traffic, not public policy. Algorithms created for the purpose of increasing ad clicks is what ends up shaping what we see online and too often, this content is not an accurate reflection of the real world. Structural flaws built into the algorithms used to sort online content may result in racial and other bias in our news feeds. As diverse voices are squeezed out, bias increases even further. This is simply not acceptable and I look forward to hearing more today about what we can do about it.

Unfortunately, forces are at work here in Washington to make this problem worse. At every turn, we see efforts to give more power to gatekeepers either by eviscerating net neutrality and privacy or by picking favorite voices for preferred regulatory treatment.

Even now, as we hold a hearing to talk about mitigating bias on the internet, FCC Chairman Pai is planning to introduce more bias into the system. The net neutrality rules that he plans to destroy are the protections that ensure that we, the

people, can decide for ourselves what we do and say online. Chairman Pai's plan will fundamentally change the free and open internet as we know it.

Independent voices—those outside the mainstream—may be most at risk simply because they don't have an affiliation with the companies that run the internet.

Unfortunately, broadband companies have more than just financial reasons to obstruct access to independent content—it can also be political. Under Chairman Pai's plan, nothing stops those in power from pushing broadband companies to censor dissenting voices or unpopular opinions or to promote views they support. We are seeing more and more often how this administration is using its political might to pressure even large companies.

This is not a partisan point or even a political one. Jeopardizing the national dialogue should concern all of us. The dialogue that happens online is critical for our democracy.

Chairman Pai's move comes after this Congress acted earlier this year to wipe out our privacy and data security online. Under President Obama, the FCC adopted fair rules to protect the little guy—ask before collecting information, don't share it without consent, and take reasonable measures to safeguard it. But that was too much for Congressional Republicans, who voted to take away these protections and hand over consumers' data to big business.

Sadly, there is still more to come. Over this past year, the FCC has taken every step possible to ensure that Sinclair Broadcasting—already the largest owner of broadcast stations in the country—becomes even bigger.

These steps by the FCC fly in the face of the laws Congress put in place to protect local voices. We understand that diverse perspectives are critical for our communities and strengthen our democracy. Instead, the FCC is doing everything it can to allow one company to control what people hear no matter where they are in the country. That is simply not what we intended.

So I look forward to discussing ways to eliminate bias in our communications systems. We need to figure out how to wrest power over information from corporations and return it back to the people.

Thank you, I yield back.

Ms. CLARKE. I thank you, Mr. Ranking Member Pallone, for yielding me time. Today's hearing is of great importance to me for various reasons, both as a congresswoman and as a consumer. You see, technology continues to touch all areas of our lives, and its reach will continue to grow in the coming days, weeks, months, and years.

With greater reach comes greater responsibility. Companies must ensure that the algorithms used for their services and products are free from all biases, including racial, ethnic, gender, sexual orientation biases. That includes making sure there is a diverse employee base behind the scenes ensuring these algorithms accurately represent American consumers.

As a member of the Congressional Black Caucus, I would like to highlight the great work of the CBC Diversity Task Force and the CBC TECH 2020 initiative, two entities that have been doing a substantive deep-dive analysis into the progress of the American tech sector in accomplishing meaningful diversity and inclusion in the technology space.

Additionally, I would like unanimous consent to submit for the record a letter my colleagues, Representatives Butterfield, Cleaver, and Kelly, and myself sent to Facebook regarding their site's use of ethnic affinity search criteria, which allow users to violate the Fair Housing Act. This is just an example of abuse within the algorithm space that really needs to be monitored and addressed, and I hope that we will get some recommendations from you here today.

It is my understanding that this is being addressed in the short term through Facebook. I just want to go on the record that this

is a concern to my colleagues and I. These issues are vitally important, and I look forward to today's testimony. Mr. Chairman, I yield back.

Mr. LATTA. And without objection, the letter is accepted for the record.

[The information appears at the conclusion of the hearing.]

Mr. LATTA. And the gentlelady yields back. This concludes the Member opening statements. The Chair reminds Members that, pursuant to the committee rules, all Members' opening statements will be made part of the record. Additionally, I ask unanimous consent that Energy and Commerce members not on the Subcommittee on Digital Commerce and Consumer Protection or the Subcommittee on Communications and Technology be permitted to participate in today's hearing. Without objection, so ordered.

Again, I want to thank our witnesses for being with us today, because it is very important for us to hear from you and being here to testify before the subcommittee. Today's witnesses will have the opportunity to give 5-minute opening statements followed by a round of questions from our Members.

Our witness panel for today's hearing will include Dr. Omri Ben-Shahar, the Leo and Eileen Herzel Professor of Law at the University Chicago of Law; Ms. Kate Klonick, the resident fellow for the Information Society Project at Yale Law School; Ms. Laura Moy, the deputy director of the Georgetown Law Center on Privacy and Technology; Dr. Catherine Tucker, the Sloane Distinguished Professor of Management and Science and Professor of Marketing at the MIT Sloane School of Management; Mr. Frank Pasquale, the Professor of Law at the University of Maryland, Francis King Carey School of Law; and Dr. Michael Kearns, the Professor and National Center Chair of the Department of Computer and Information Science at the University of Pennsylvania.

Again I want to thank all of our witnesses for being with us today, and again you each have 5 minutes. If you will, just pull that mic up close and turn on the button. We look forward to hearing your testimony.

And Doctor, we will start with you this morning. Thank you.

STATEMENTS OF OMRI BEN-SHAHAR, PH.D., LEO HERZEL PROFESSOR IN LAW, UNIVERSITY OF CHICAGO LAW SCHOOL; KATE KLONICK, RESIDENT FELLOW, INFORMATION SOCIETY PROJECT, YALE LAW SCHOOL; LAURA MOY, DEPUTY DIRECTOR, CENTER ON PRIVACY & TECHNOLOGY AT GEORGETOWN LAW; CATHERINE TUCKER, PH.D., SLOANE DISTINGUISHED PROFESSOR OF MANAGEMENT SCIENCE, MIT SLOANE SCHOOL OF MANAGEMENT; FRANK PASQUALE, PROFESSOR OF LAW, UNIVERSITY OF MARYLAND; AND, MICHAEL KEARNS, PH.D., COMPUTER AND INFORMATION SCIENCE PROFESSOR, UNIVERSITY OF PENNSYLVANIA

STATEMENT OF OMRI BEN-SHAHAR

Dr. BEN-SHAHAR. Thank you, Chairman Latta. Thank you, Chairman Blackburn, for inviting me, Ranking Members Schakowsky and Doyle and members of the subcommittee, I cherish this opportunity to participate in the conversation.

I am a law professor at the University of Chicago, and I specialize in consumer law and consumer protection. You will hear today a lot about the dangers of big data enterprise, how websites know our locations, how smart alarms know and predict our vacations, how employers and insurers know our medications, and even Fitbit records our dedication.

We of course all know the data-driven economy delivers enormous convenience and benefits too by offering personalized experience to consumers, but concerns about discrimination, manipulation, data security, and market power and the potential harms they might cause ought to be taken seriously. Still, it is important throughout this inquiry that the basic question—What is the consumer injury?—be answered before we begin thinking about what the solution ought to be.

You will probably hear today other speakers call for more transparency on how data is used and secured so as to give consumers more control over their data and allow them to make more informed decisions. Chairman Walden invited such noble proposals of transparency, writing eloquently in an op-ed, quote, “It is our job to shine the light on these practices for consumers and ensure transparency in the marketplace so that they can make informed choices.”

I would like to spend my remaining 4 minutes or so to try to talk you out of this transparency instinct. It is not that I don’t like transparency or informed decision, it is just that this technique has never worked in any area, and it is decisively unlikely to yield any benefit here. I co-authored a book titled “More Than You Wanted To Know,” in which I looked at the effect of transparency laws. These are the numerous laws that require companies to give consumers full disclosures to help consumers make informed choices.

Mandated disclosure is probably the most common and for sure the least successful regulatory technique in American law. Disclosure requirements, we sometimes call them sunshine laws, have been used for decades as the primary tool for consumer protection to protect borrowers, investors, medical patients, internet users, insurance buyers, home buyers, in every area of the law, and the record confirmed by mountains of empirical evidence is abysmal—transparency doesn’t make a difference.

Transparency requires that companies give consumers disclosures, but consumers are not cooperating. They are not reading or using the disclosures. How could they? The texts are too long and cluttered.

[Photo shown.]

Here is a picture of a typical artifact of transparency, Apple’s terms and conditions that include their privacy policy, which I printed out and assembled into a 30-foot scroll, 8-point font, mind you, and hung from the top of the atrium at the University of Chicago Law School.

Shoving this monstrosity in front of consumers: Is that what consumer protection ought to do? If consumers tried to read the disclosures, they would of course not understand them and would not be able to put them to profitable use. To use complex information, one needs experience and expertise which people simply do not have.

Transparency is defeated not because it is a bad idea but because it is so overused.

When you close a mortgage, you receive at least 50 different disclosures so that you, quote, “know before you owe.” When you walk into a clinic or buy a product or enter a website or download an app or eat at a restaurant or check your bank balance, you receive disclosures, all in the name of transparency. Consumers have long become numb and indifferent.

Any transparency effort in the area of data protection would meet the same consumer apathy. Do you really want to be the authors of an irrelevant policy? Can transparency be done more effectively? If disclosures are defeated by complexity, can simplicity save them? Simplification seems like an obvious solution: If disclosures are too long, shorten them; if too technical, use plain language; if poorly presented, improve the formatting. Unfortunately, simplification strategies have been tried for as long as disclosures have failed.

In my research, I tested whether people who are sharing deeply private information with websites that engage in nasty data practices can be prompted to act more prudently by well-designed privacy warnings. I discovered that no matter how simple, conspicuous, and alarming the warning the consumers receive, their behavior is entirely unchanged. Consumers don't pay attention to any of the transparency tools lavished upon them.

To conclude, if Members of Congress believe that collection of consumers' data poses risks that require regulatory intervention, I advise that they look for solutions that are outside the popular but unsuccessful repertoire of mandated disclosure and transparency. Thank you.

[The prepared statement of Dr. Ben-Shahar follows:]



1111 East 60th Street | Chicago, Illinois 60637
 phone 773-702-2087 | fax 773-702-0730
 e-mail omri@uchicago.edu
home.uchicago.edu/omri

Omri Ben-Shahar
Leo Herzel Professor in Law
Director, Coase-Sandor Institute for Law and Economics

The Failure of Transparency

Testimony of Professor Omri Ben-Shahar, University of Chicago
 Before
 Committee on Energy and Commerce
 Subcommittee on Communications and Technology
 Subcommittee on Digital Commerce and Consumer Protection

Introduction and Summary

The massive collection of people's personal information by companies impacts consumers' privacy and security. At present, the primary and almost exclusive way in which consumers are protected is through "transparency": requiring that companies disclose to consumers what information they collect and how they use it, and alert consumers in the event of a data security breach.

I am a professor of law at the University of Chicago, specializing in consumer markets. I have studied the effects of mandated disclosures in the area of data privacy and in every other area of consumer protection. My research, summarized in a recent book titled "More Than You Wanted To Know" (Princeton, 2014), concludes that disclosure rules are entirely ineffective.

Transparency is intended to strengthen competition in the market. Mandated disclosures are aimed at helping consumers make informed choices and inducing companies to act honestly. It was Louis Brandeis who, 100 years ago, said "sunlight" is "the best of disinfectants." But disclosure rules have miserably failed to achieve their goals. Massive amounts of evidence show that people don't read the disclosures and don't use them to make more informed choices. In reality, disclosures are regularly ignored. They are an empty ritual.

It is tempting to think that disclosures can be more effective if designed to deliver information to consumers in simpler formats. But simplification, too, has been tried for decades and failed. My research shows that simplified disclosures about data privacy and security will have no effect on the behavior of consumers or the companies that collect their information.

Thus, if members of Congress believe that the collection of consumers' data poses risks that require legal intervention, I advise that they look for regulatory solutions that are outside the popular but unsuccessful repertoire of disclosure and transparency.

Disclosure is the Primary Protection Under the Law

The collection of consumers' personal information by companies poses two fundamental challenges. The first is privacy: much of the information collected is personal and sensitive. The second is security: the information may be hacked or stolen and then used in ways detrimental to consumers' financial safety.

American law imposes few practical limits on the collection of personal information by companies. It also does not establish concrete standards for data protection and security. Instead, the most common protection for privacy and data security is "transparency": that any collection or security breach of personal data be accompanied by full and conspicuous disclosures to consumers. Much of the attention of lawmakers, judges, and commentators is directed to "shine the light"—to guarantee that full disclosures are in place to help consumers make more informed and safe choices.

Unfortunately, disclosure regulation has largely failed. And there is little reason to hope that it will ever succeed. Disclosures' failures have a long and persistent history, occurring without exception in every domain of consumer protection. The evidence of failure is abundant, and it is largely uncontested in the literature.

Mandated disclosure is the primary tool of data privacy protection. Our legal environment is packed with statutes and regulations that prohibit various types of data collection or surveillance, but almost all such prohibitions may be waived by consumers. If the consumer agrees, almost any personal information may be collected. It is exceedingly easy for companies to get consumers to agree to waive the statutory protections. It only takes a click "I agree" to the "terms and conditions" or the "privacy notice" (legal texts that regularly contains thousands of words). In fact, a click is not even necessary—the requirement of "informed consent" is satisfied if companies prominently post their privacy notices" on their webpages. Because companies largely comply with the disclosure requirements, the great majority of courts are finding that consumers are effectively agreeing to the data collection, rendering it perfectly legal.

American law also imposes few specific regulations on the security and protection of consumers' information. Businesses are encouraged by the FTC to engage in "best practices" in the storage and safeguarding of consumers' data, but the most concrete obligation is, again, disclosure. For example, under California law businesses are required to "disclose the breach of the security . . . in the most expedient time possible and without unreasonable delay." (California Civil Code, Sec. 1798.82).

The Allure of Disclosure

Transparency is the most common technique of consumer protection not only in the area of privacy and data security. It is the primary tool for protecting borrowers, investors, medical patients, homebuyers, insurance policy holders, internet users—every sector with its array of mandated disclosures. In each of these areas, people are

making choices that are often complex and could severely impact their well being, and are doing so without being fully aware of the risks and benefits. The solution seems alluringly simple: if people make poor decisions because they have poor information, give them more information! Don't people want to make decisions for themselves, and to make them well? Isn't more information better than less? Wouldn't people gratefully take and earnestly use information they are offered?

Because it is so sensible, and because it is thought to be at worst harmless, the mandated disclosure technique is a political winner. Disclosure laws have no enemies as they resonate with almost all American ideologies. Disclosure laws appeal to free-market proponents and to progressives alike, to Democrats and Republicans, even to budget hawks. In almost every area, disclosure mandates and "sunshine laws" are enacted with almost no opposition. Even business interests acquiesce, as they prefer disclosure mandates to more intrusive command-and-control regulations.

Mandated disclosure is alluring because its failures are little noticed and soothingly explained. Lawmakers and commentators do not realize that it is a method so extensively tried, and so they readily attribute any documented failure to the particular way the disclosure was implemented. Maybe the disclosure failed because it was too narrow, or maybe too broad. Maybe it failed because it was too short, or maybe too long. Maybe it was recited to the consumer prematurely, or maybe it was given too late in the game. Maybe it was too technical. Excuses abound.

The Failure of Disclosure

Mandated disclosure is alluring, but it routinely fails to achieve its ambitious goals. Empirical studies show that disclosures rarely change the decisions that people make. People don't read the disclosures. If they read, they do not understand the texts, often written at superior levels of literacy. And even if the texts are written in lay language, they cannot use them profitably because the information conveyed is complex and using complex information to make good decisions requires experience and expertise.

The problem with mandated disclosures is not just their length and complexity. It is also their accumulation. Because disclosures have been enacted in so many areas for so long, people are swamped with disclosures, notices, and warnings of all types. Consumers have become numb to these rituals, viewing them as annoying "fine print" that can be safely ignored. Consumers' apathy is entirely rational: there is simply not enough time to review all the disclosures that the law requires companies to bestow upon them. Just to read all the privacy notices a typical person receives every year would take—according to an estimate done a decade ago—76 days of full time reading, with the loss of productive time costing the economy \$781 billion. And, recall, privacy notices are only a small fraction of the sum total of mandated disclosures consumers receive.

Much evidence shows that the disclosure of information is almost irrelevant. Consumers ignore mortgage and banking disclosure (how could they not, given the length and complexity of such documents?) Warnings about product risks or conflicts of interests, medical consent forms, even food and nutrition labels—are all falling upon deaf ears. The evidence is strikingly disappointing: “transparency” requirements have not improved the market outcomes for consumers in any meaningful way.

The Failure of Simplification

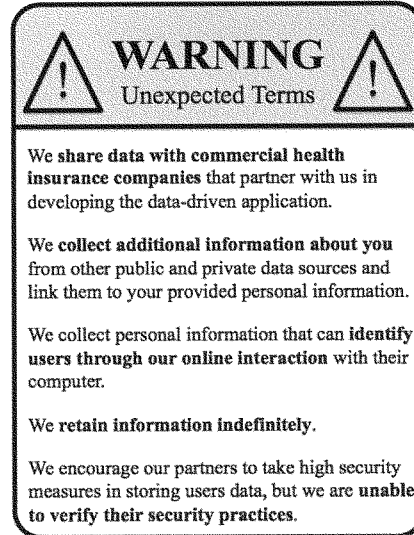
Can disclosure be done more effectively? If existing disclosures are defeated by complexity, can simplicity save them? Could the formats of disclosure be reengineered to become more accessible to consumers and more effective? Simplification seems like an obvious solution. If a disclosure is too long, shorten it. If it's too technical, make it more user-friendly. If it's poorly presented, improve the formatting.

Simplification faces a paradox: if you give people less than full information they may overlook some risks; but if you give them full information they might find the disclosure too long and unmanageable. The pragmatic solution is to focus consumers' attention to the most crucial information and present it in non-technical, easily comparable, format.

Many kinds of simplification strategies along this line have been tried in various areas, with little or no success. Even the simplest of all disclosures—the truth-in-lending's “APR” score that has to be presented before consumers take a loan—has demonstrably failed to improve borrowing decisions. More recent “behaviorally-informed” disclosures, designed by social scientists schooled in diagnosing peoples' decisional failures, have similarly yielded deeply disappointing effects. Such “smart disclosure” designs are required, for example, under federal credit card laws, but have generated only small, almost microscopic, impact on consumers' behavior.

In my own research, I have examined various formats of simplified privacy disclosures. I tested whether people who engage in activity that raises heightened privacy concerns are prompted to act prudently when shown well-designed privacy warnings. I discovered that no matter how simple and conspicuous the warning, consumers' behavior is unchanged. It doesn't matter if the privacy warning is cluttered (as many currently are), or instead drafted according to the FTC's “best practices” guidelines. It doesn't even matter if the privacy notice is pared down to a simple warning box, similar to the familiar Nutrition Facts box (see image below). The simplification of the disclosure has no effect. Consumers don't read the warning one way or another, and imprudently share the same amount of personal information, regardless of the disclosure's format.

Simplification is failing, and this should not be a big surprise. Is it really possible to simplify the complex? Disclosures are long and tiresome because the information necessary to make good decisions about unfamiliar issues is complex and nuanced.



“Warning Label” Privacy Disclosure

Beyond Disclosure

I wrote a book about the failure of disclosure titled “More Than You Wanted to Know.” I presented my findings and conclusions—that mandated disclosure fails and cannot be fixed—to numerous audiences. Most agree with my claims, because they know from their own experience that they, too, don’t read and are not being helped by disclosures. Still, at the end I am always asked, “What, then? If not disclosure, what does work?”

Unfortunately, there is no one-size-fits-all solution, no new panacea. Different problems merit different solutions. In the area of data privacy and security, it is necessary to begin by identifying the harm from which consumers have to be protected. Collection of information by companies is not harmful in itself. The great majority of consumers are happy to pay for excellent services with their data rather than with money. Some research shows that consumers are not willing to pay more than a few dollars to prevent the harvesting of their data by websites they visit or apps they use. The various class action lawsuits that allege privacy violations have so far failed to robustly demonstrate actual concrete injuries. Moreover, markets seem to be providing some protection: companies that collect sensitive information implement great safeguards. Adult websites, for example, are far more restrictive than other platforms about data sharing; and cloud storage services have higher data security standards. Data security breaches are not harmful unless the data is used for fraudulent transaction. A legal scheme insuring consumers against such losses may be necessary to the extent that consumers are not already protected or insured.

Mr. Latta. Thank you very much for your testimony this morning.

And, Ms. Klonick, you are recognized for 5 minutes.

STATEMENT OF KATE KLONICK

Ms. Klonick. Thank you. Chairmen Blackburn and Latta, Ranking Members Doyle and Schakowsky, and members of the subcommittees, thank you for having me here to discuss this important topic.

Every day millions of people around the world post videos, pictures, and text to online speech platforms, but not everything that is posted remains there. Sites like Facebook, Twitter, and YouTube actively curate the content that is posted by their users through a mix of algorithmic and human processes broadly termed content moderation. Until recently, how and why these platforms made these decisions on users' speech was largely opaque.

Over the last 2 years, I have interviewed dozens of former and current executives at these platforms as well as content moderation workers at these companies working abroad in an effort to better understand how and why these platforms regulate content. A summary of that research and my conclusions are the subject of my paper, "The New Governors: The People, Rules, and Processes Governing Online Speech," forthcoming in the Harvard Law Review. My testimony today draws from that expertise and knowledge that I gained in researching and writing that article.

As a threshold matter, when I refer to content moderation I am referring specifically and exclusively to the experience of the user in posting speech to a platform and what happens to that posted content in terms of removal or nonremoval. I am not speaking to the algorithm that configures the prioritization, promotion, order, or frequency of how content later appears in users' news feeds or Twitter feeds.

And in that context, content moderation happens at many levels. It can happen before content is actually published on the site, and when a user uploads a photo, a message appears: "Upload completed. The video in your post is being processed. We will send you a note when it is ready for review." And the moderation process that happens in this moment between upload and publication largely runs through an algorithm screening that checks for matches in pixel fingerprints between illegal or banned content and the uploaded content. Examples of this include photo DNA for child pornography and content ID for copyrighted information.

Only a very small amount of material is removed through these types of processes, and most is published, and once published it can be removed in two ways. The first is by platforms proactively using their own moderators, but because of the absolutely enormous amount of posts, this is not a feasible method for all but a very select area of moderation, such as extremist and terrorist content.

The second way content is removed after publication is also how the vast majority of content is removed, through being flagged as violating community standards by other users on the site. After a piece of content is flagged, it will stay up, but a crop screen grab of the content is placed in a database queue, where it is eventually

reviewed by trained human decision makers. They will look at the offending content and see if it actually violates the terms of service.

With that background, I would like to use my brief time to clarify four major misconceptions about content moderation. First, that, contrary to this hearing's title, the vast majority of content moderation of user content is done by trained human decision makers who review content only after it has been flagged by other users and not by algorithms or AI or photo recognition.

Second, while users who use sites like Facebook are given a public set of community standards guiding what kinds of content is posted by the site, a separate and much more detailed and much more regularly updated set of internal rules is used by human moderators in making their decisions. These internal rules at these companies are not currently known to the public.

Third, Facebook and most platforms use one global set of rules with exceptions to comply with the laws of a given jurisdiction to curate content. This means, for example, the definitions of inappropriate sexual activity are the same for users in Canada as they are for users in India as they are for users in France.

Finally, it is critical to note that the ability for these platforms to create this intricate system of governance to regulate content stems from incentives put in place by Communications Decency Act Section 230 which granted platforms immunity from intermediary liability in an effort to encourage sites to remove offensive content while also protecting against collateral censorship of users' speech.

In many ways these platforms' self-regulation have very well met the goals of Section 230, but as access to online speech platforms has increasingly become an essential public right, new concerns about regulating platforms are being raised. While these and other concerns are undoubtedly present, changes to Section 230 or new regulations that might affect it should be considered with extreme caution and with a full appreciation of the potential damage that could be caused to consumer rights and to free speech. Thank you.

[The prepared statement of Ms. Klonick follows:]



Yale Information Society Project

**House of Representatives Committee on Energy and Commerce
Subcommittee on Communications and Technology**

“Algorithms: How Companies’ Decisions About Data and Content Impact Consumers”
November 29, 2017

Written Remarks of Kate Klonick
Yale Law School Information Society Project*

Chairman Blackburn, Ranking Member Doyle, and Members of the Subcommittee:
Every day millions of people around the world post pictures, videos, and text to online speech platforms, but not everything that is posted remains there. Sites like Facebook, Twitter, and YouTube actively curate the content that is posted by their users through a mix of algorithmic and human processes, broadly termed content moderation. Until recently, how and why these platforms made these decisions on user speech was largely opaque. For two years I have interviewed over three-dozen former and current executives and content moderation workers at these companies in an effort to better understand how and why these platforms regulate content.

This written testimony borrows heavily from my Article summarizing those findings¹ and attempts to clarify a few major points about content moderation, including:

- The vast majority of content moderation of user content (roughly estimated at over 90%) is done by trained human content moderators who review content only after it has been flagged by platform users and **not by algorithms**, contrary to this hearing’s title.
- While users at sites like Facebook are given a public set of “Community Standards” guiding what kind of content is posted on the site, a separate much more detailed, and much more regularly updated set of internal rules is used by human moderators in making their decisions. These internal rules, at least at Facebook, are not currently known to the public.²
- That Facebook, and most platforms, use one global set of rules (with exceptions to comply with Nation-State laws) to curate content. This means, for example, that definitions of “inappropriate sexual activity” are the same for users in Canada, as they

* Resident Fellow at the Information Society at Yale Law School; Ph.D Candidate in Law, Yale University; J.D. Georgetown University Law Center; A.B. Brown University. I’m testifying on own behalf, not on behalf of my employer or anyone else.

Email: kate.klonick@yale.edu Website: www.kateklonick.com

¹ Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, forthcoming HARV. L. REV. (2018). Available for download at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2937985

² In May 2017, *The Guardian* published a series of documents claiming to be the “leaked rules” of Facebook. In fact, these were not the precise rules, but rather slides used to train human content moderators on Facebook’s internal rules. Nick Hopkins, *Revealed: Facebook’s internal rulebook on sex, terrorism and violence*, GUARDIAN (May 21, 2017) <https://www.theguardian.com/news/2017/may/21/revealed-facebook-internal-rulebook-sex-terrorism-violence>.



Yale Information Society Project

are for users in India, as they are for users in France—irrespective of the norms of each country.

- These platforms intricate systems of governance to regulate content are a response to the Communications Decency Act Section 230, which incentivized sites to remove offensive content with immunity from intermediary liability.³ In many ways, these platforms' self-regulation has met the goals of Section 230, but as access to online speech platforms has increasingly become an essential public right⁴ new concerns about the expansive immunity granted under Section 230 are being raised. While these and other concerns are undoubtedly present, changes to Section 230 or new regulation that might affect it, should be considered with extreme caution and with a full appreciation of the potential damage that could be caused to consumer rights.
- While there have long been worries about internet service providers favoring access to some content over others, there has been less concern about companies further along the pipeline holding an internet on/off switch. In large part, this is because at other points in the pipeline, users have choice. But the fewer choices you have for the infrastructure you need to stay online, the more serious the consequences when companies refuse service. This is one important reason net neutrality is so important. As Section 230 reveals, we generally agree that it's appropriate for social media companies to take down certain kinds of content — that's how they ensure our newsfeeds aren't full of pornography or violence. But that doesn't mean we don't want that type of content to be able to exist *somewhere* on the Internet. Ensuring that ISPs remain neutral is necessary to guaranteeing the continuation of a free and open Internet.

How Platforms Moderate Content

Content moderation happens at many levels. It can happen before content is actually published on the site as with *ex ante* moderation, or after content is published, as in *ex post* moderation. These methods can be either *reactive*, in which moderators passively assess content

³ The ability of private platforms to moderate content comes from § 230 of the Communications Decency Act, which gives online intermediaries broad immunity from liability for user generated content posted on its site. 47 U.S.C. § 230. The purpose of this grant of immunity was both to encourage platforms to be “Good Samaritans” and take an active role in removing offensive content, and also to avoid free speech problems of collateral censorship. See *Zeran v. Am. Online, Inc.* 129 F.3d 327, 330 (4th Cir. 1997) (discussing the purposes of intermediary immunity § 230 were not only to incentivize platforms to remove indecent content, but to protect the free speech of platform users). See also Eric Goldman, *Ten Worst Section 230 Rulings of 2016 (Plus the Five Best)*, (Jan. 4, 2017) at <http://blog.ericgoldman.org/archives/2017/01/ten-worst-section-230-rulings-of-2016-plus-the-five-best.htm>. For a comprehensive and complete cataloging of § 230 cases with context and commentary, see Professor Eric Goldman's blog, <http://blog.ericgoldman.org/>.

⁴ *Packingham v. North Carolina*, 137 S.Ct. 1730 (2017) (holding that a state statute barring registered sex offenders from using online social media platforms was unconstitutional under the First Amendment). In his opinion, Justice Kennedy wrote that “[w]hile in the past there may have been difficult in identify the most important places (in a spatial sense) for the exchange of views, today the answer is clear. It is cyberspace—the ‘vast democratic forums of the Internet’ in general, and social media in particular.” *Id.* at 1735 (quoting *Reno v. ACLU*, 521 U.S. 844, 868 (1977)).



Yale Information Society Project

and update software only after other users bring the content to their attention, and *proactive* moderation, in which teams of moderators actively seek out published content for removal. Additionally, these processes can be *automatically* made by software or algorithms, or *manually* made by humans.⁵

1. *Ex Ante* Content Moderation⁶

When a user uploads a video to Facebook, a message appears: “Upload Completed: The video in your post is being processed. We’ll send you a notification when it’s done and your post is ready to view.”⁷ *Ex ante* content moderation is the process that happens in this moment between upload and publication. The vast majority of *ex ante* content moderation is an automatic process largely run through algorithmic screening without the active use of human decision-making.

An example of such content is child pornography, which can reliably be identified on upload to a site through a picture recognition algorithm called PhotoDNA.⁸ Under federal law, production, distribution, reception, and possession of an image of child pornography is illegal, and as such, sites are obligated to remove it.⁹ A known universe of child pornography—around 720,000 illegal images—exists online.¹⁰ By converting each of these images to gray scale formatting,

⁵ Cf. James Grimmelmann, *The Virtues of Moderation*, 17 YALE J.L. & TECH. 42, 63-70 (2015) (describing how a moderation system operates through distinctions between automatic, manual, transparent, secret, *ex ante*, *ex post*, centralized, and decentralized features). Grimmelmann’s taxonomy, while foundational, speaks more generally to all of Internet moderation rather than content publishing platforms, specifically. In the context of speech, the distinction between *ex ante* and *ex post* is especially important to determine if moderation is happening before or after publication. Of secondary concern is whether content is being moderated through reaction or through proactive measures. Finally, for the purposes of this hearing, the distinction between automatic or algorithmic moderation and human manual moderation is of central importance.

⁶ Because it happens before publication takes place, *ex ante* content moderation is the type of prior restraint that scholars like Professor Jack Balkin are concerned with. See Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296, 2299 (2014). Of the two automatic means of reviewing and censoring content—algorithm or geo-blocking—geo-blocking is of more concern for the purposes of collateral censorship and prior restraint. In contrast, algorithm take down is currently used to remove illegal content like child pornography or copyright violations. *But see* Rebecca Tushnet, *Power Without Responsibility: Intermediaries and the First Amendment*, 76 GEO. WASH. L. REV. 986, 1003-05 (2008) (noting that the DMCA notice-takedown provisions give platforms no incentive to investigate and therefore “suppress critical speech as well as copyright infringement.”).

⁷ FACEBOOK, UPLOADING & VIEWING VIDEOS (accessed Mar. 1, 2017) https://www.facebook.com/help/154271141375595/?helpref=hc_fnav

⁸ Tracy Ith, *Microsoft’s PhotoDNA: Protecting children and businesses in the cloud*, MICROSOFT NEWS (accessed Mar. 1, 2017) <https://news.microsoft.com/features/microsofts-photodna-protecting-children-and-businesses-in-the-cloud/#sm.001eom8zb14bad5html1ixrkpzssa>.

⁹ See 18 U.S.C. § 2251; 18 U.S.C. § 2252; 18 U.S.C. § 2252A. It is important to remember that § 230 expressly states that no Internet entity has immunity from federal criminal law, intellectual property law or communications privacy law. This means that every Internet service provider, search engine, social networking platform and website is subject to thousands of laws, including child pornography laws, obscenity laws, stalking laws and copyright laws. 47 U.S.C. § 230 (e).

¹⁰ This “known universe” of child pornography is maintained and updated by the International Centre for Missing and Exploited Children and the U.S. Department of Homeland Security in a program known



Yale Information Society Project

overlaying a grid, and assigning a numerical value to each square, researchers were able to create a “hash” or signature that remained even if the images were altered. As a result, platforms can determine within micro-seconds between upload and publication if an image contains child pornography.¹¹ Geo-blocking is another form of automatic *ex ante* moderation. Unlike PhotoDNA, which prevents the publication of illegal content, geo-blocking prevents both the publication and viewing of certain content based on a user’s location. As happened in the controversy over the *Innocence of Muslim* video, geo-blocking usually comes at the request of a government notifying a platform that a certain type of posted content violates its local laws.

It is important to note that, of course, algorithms do not decide for themselves which kind of content they should block from being posted. Content screened automatically is typically content that can reliably be identified by software and is illegal or otherwise prohibited on the platform. This universe of automatically moderated *ex ante* content is regularly evaluated and updated through iterative software updates and machine learning. For example, in a similar fashion to PhotoDNA, potential copyright violations can be moderated proactively through software like ContentID. Developed by YouTube, ContentID allows creators to give their content a “digital fingerprint” so it can be compared against other uploaded content. Copyright holders can also flag already published copyright violations through notice and takedown.¹² These two systems work together, with user-flagged copyrighted material eventually added to ContentID databases for future proactive review. This mix of proactive, manual moderation, informed and automatic *ex ante* moderation is also evident in the control of spam. All three platforms (and most Internet companies, generally) struggle to control spam postings on their sites. Today, spam is mostly blocked automatically from publication through software. Facebook, Twitter, and YouTube, however, all feature mechanisms for users to report spam manually.¹³ *Ex ante* screen software is iteratively updated to reflect these flagged spam sources.

2. *Ex Post* Proactive Manual Content Moderation

Recently, a form of content moderation that harkens to the earlier era of AOL chat rooms has re-emerged: platforms proactively using their own moderators, instead of relying on flagging by users to seek out and remove published content. Currently, this method is largely confined to the moderation of extremist and terrorist speech. As of February 2016, dedicated teams at Facebook proactively removed all posts or profiles with links to terrorist activity.¹⁴ Such efforts

as Project Vic. Mark Ward, *Cloud-based archive tool to help catch child abusers*, BBC NEWS (Mar. 24, 2014) <http://www.bbc.com/news/technology-26612059>.

¹¹ *Ith*, *supra* note 8.

¹² See e.g., YOUTUBE, *YouTube Help: Submit a copyright takedown notice*, <https://support.google.com/youtube/answer/2807622> (last visited Aug. 15, 2016).

¹³ See e.g. Panda Security, *How Twitter aims to prevent your timeline from filling up with spam* (Sept. 12, 2014) <http://www.pandasecurity.com/mediacenter/social-media/twitter-spam/>; James Parsons, *Facebook’s War Continues Against Fake Profiles and Bots*, HUFF. POST (May 22, 2015) http://www.huffingtonpost.com/james-parsons/facebooks-war-continues-against-fake-profiles-and-bots_b_6914282.html.

¹⁴ Natalie Andrews & Deepa Seetharaman, *Facebook Steps Up Efforts Against Terrorism*, WALL ST. J. (Feb. 11, 2016), <http://www.wsj.com/articles/facebook-steps-up-efforts-against-terrorism-1455237595>.



Yale Information Society Project

were doubled in the wake of terrorist attacks and the events in Charlottesville.¹⁵ This is an important new development affecting content moderation with an ever-evolving balance between ensuring national security yet maintaining individual liberty and freedom of expression, but it still only comprises a small amount of the total moderation that happens on these sites.

3. *Ex Post* Reactive Manual Content Moderation

As previously mentioned, with the exception of proactive moderation for terrorism described above, almost all user-generated content that is published is reviewed *reactively*, that is, through *ex post* flagging by other users and reviewed by human content moderators against internal guidelines. Flagging—alternatively called reporting—is the mechanism provided by platforms to allow users to express concerns about potentially offensive content.¹⁶ The adoption by social media platforms of a flagging system serves two main functions: (1) it is a “practical” means of reviewing huge volumes of content, and (2) its utilization of users serves to legitimize the system when platforms are questioned for censoring or banning content.¹⁷

Facebook users flag over one million pieces of content worldwide every day.¹⁸ Content can be flagged for a variety of reasons and the vast majority of items flagged do not violate the Community Standards of Facebook. Instead they often reflect internal group conflicts or disagreements of opinion. To resolve the issue, Facebook created a new reporting “flow”—the industry term to describe the sequence of screens users would experience as they made selections—that would encourage users to resolve issues themselves rather than report them for review to Facebook.¹⁹ Facebook has also designed its reporting flow to triage flagged content for review. This makes it possible for Facebook to immediately prioritize certain content for review, and when necessary, notify authorities of emergency situations like suicide, imminent threats of violence, terrorism, or self-harm. Other content, like possible hate speech or harassment, can be queued into less urgent databases for general review.²⁰

When content is flagged or reported it is sent to a server where it awaits review by a human content moderator. At Facebook, there are three basic tiers of content moderators: “Tier 3” moderators, who do the majority of the day-to-day reviewing of content; “Tier 2” moderators, who supervise Tier 3 moderators and review prioritized or escalated content; and “Tier 1” moderators, who are typically lawyers or policy makers based at company headquarters.

¹⁵ *Id.*

¹⁶ Kate Crawford & Tarleton Gillespie, *What is a flag for? Social media reporting tools and the vocabulary of complaint*, NEW MEDIA & SOC. (2014), at 2.

¹⁷ *Id.* at 3.

¹⁸ See Catherine Buni & Soraya Chemaly, *The Secret Rules of the Internet*, THE VERGE (Mar. 13, 2014), www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech.

¹⁹ *Radiolab: The Trust Engineers*, WNYC (Feb. 9, 2015) (downloaded using iTunes).

²⁰ *Facebook Reporting Guide: What Happens When You Report Something?*, uploaded to Scribd June 19, 2012 by Facebook Washington DC, <https://www.scribd.com/doc/97568769/Facebook-Reporting-Guide>. After content has been flagged to a platform for review, the precise mechanics of the decision-making process become murky. Platforms do not publish details of their internal content moderation guidelines; no major platform has made such guidelines public. Buni & Chemaly, *supra* note 18.



Yale Information Society Project

In the early days—before 2008 to 2009—recent college graduates based in the San Francisco Bay Area did much of the Tier 3 content moderation.²¹ Today, most platforms, including Facebook, either directly employ content moderation teams or outsource much of their content moderation work to companies based in the Philippines, Ireland, Singapore, India, or Eastern Europe.²² Today, Tier 3 moderators typically work in “call-centers” in the Philippines, Ireland, Singapore, India, or Eastern Europe. Within Facebook, these workers are called “community support” or “user support teams.”²³

Tier 2 moderators are typically supervisors of Tier 3 moderators or specialized moderators with experience judging content. They work both remotely (many live in the United States and supervise groups that are internationally based) and locally at call-centers.²⁴ Tier 2 moderators review content that has been prioritized, like imminent threats of violence, self-harm, terrorism, or suicide that arrive to Tier 2 directly through the reporting flow or are identified and escalated to Tier 2 by Tier 3 moderators. Tier 1 moderation is predominantly performed by the legal or policy headquarters of a platform. At Facebook, for example, a Tier 3 worker could be based in Hyderabad, the Tier 2 supervisor could be based in Hyderabad, or remotely in a place like Dublin, but a Tier 1 contact would be based in Austin, Texas or the San Francisco Bay Area.

At Facebook, Tier 3 moderators have three decision-making options regarding content: they can “confirm” the content violates the Community Standards and remove it, “unconfirm” that the content violates Standards and leave it up, or escalate review of the content to a Tier 2 moderator or supervisor. The internal rules describe certain types of content requiring mandatory escalations. For example in 2012 at Facebook: child nudity or pornography, promotion or encouragement of bestiality, credible threats, bullying, self-harm content, poaching of endangered animals, Holocaust denial, all attacks on Ataturk, maps of Kurdistan and Burning Turkish Flags.²⁵ If a moderator has decided to ban content, a Facebook user’s content is taken down, and she is automatically signed off of Facebook. When the user next attempts to sign in, she will be given the following message explaining without detail that an offensive post was removed in violation of community standards. At Facebook, users who repeatedly have content removed are gradually escalated in punishment: two removed posts in a certain amount of time, for example, might mean your account is suspended for 24-hours.

²¹ Buni & Chemaly, *supra* note 18.

²² Adrian Chen, *The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed*, Wired, (Oct. 23, 2014) <https://www.wired.com/2014/10/content-moderation/>; Adrian Chen, *Inside Facebook’s Outsourced Anti-Porn and Gore Brigade, Where ‘Camel Toes’ are More Offensive Than ‘Crushed Heads.’* GAWKER (Feb. 16, 2012) <http://gawker.com/5885714/inside-facebooks-outsourced-anti-porn-and-gore-brigade-where-camel-toes-are-more-offensive-than-crushed-heads>. Within Facebook, these workers are called “community support” or “user support teams.”

²³ *Id.*

²⁴ *Id.*; Telephone Interview with Dave and Charlotte Willner (Mar. 23, 2016).

²⁵ Abuse Standards (AS) 6.1 available at <https://www.scribd.com/doc/81863464/oDeskStandards>; Abuse Standards (AS) 6.2 available at <https://www.scribd.com/doc/81877124/Abuse-Standards-6-2-Operation-Manual-hereinafter-collectively-Abuse-Standards>. These are copies of documents that were leaked from a content moderator working at oDesk (now UpWork) doing content moderation for Facebook. They are not the actual internal rules of Facebook, but they were oDesk’s approximation of Facebook’s rules in 2012.



Yale Information Society Project

Normative Implications of Platform Governance on Potential Regulation

These details about how and why platforms are governing user speech have direct implications on potential regulation and our understanding of online speech.

1. Any reform to Section 230 should be approached with caution

When CDA Section 230 was put into place in 1996, the Internet was a very different place. Spam and pornography were threatening to dominate platforms, but courts were beginning to hold platforms civilly liable if they acted to remove such content.²⁶ Section 230 lifted the “specter of tort liability” that might “deter service providers from blocking and screening offensive material” and also result in platforms removing too much user speech resulting in an “obvious chilling effect.”²⁷ “Faced with potential liability for each message republished by their services, interactive computer service providers might choose to severely restrict the number and type of messages posted.”²⁸

In many ways, these major social media platforms’ self-regulation has met the goals of Section 230—removing content that users find normatively unpalatable, while keeping up as much content as possible.²⁹ But in the 21 years since Section 230 was passed, access to online speech platforms has increasingly become an essential public right new and concerns about the expansive immunity granted under Section 230 are being raised. While these and other concerns are undoubtedly present, changes to Section 230 or new regulation that might affect it, should be considered with extreme caution and with a full appreciation of the potential damage that could be caused to consumer rights and free speech online.

2. Speech platforms’ ability to self-regulate content has little to no direct applicability to broadband ISPs ability to self-regulate

Generally speaking, there are two kinds of corporate players on the internet: companies that build infrastructure through which content flows, and companies that seek to curate content and create a community. Internet service providers like Verizon and Comcast, domain name servers, web hosts and security services providers are all the former — or the “pipe.” They typically don’t look at the content their clients and customers are putting up, they just give them the means to do it and let it flow. Social media platforms like Facebook are the latter. They encourage their users to create, share and engage with content — so they look at content all the time and decide whether they want to allow hateful material like that of neo-Nazis to stay up.

²⁶ See *Cubby v. CompuServe*, 776 F. Supp. 135, 138 (S.D.N.Y. 1991) (holding CompuServe could not be held liable for the defamatory content because the intermediary did not review any of the content posted to the forum) and *Stratton Oakmont, Inc. v. Prodigy Servs. Co.*, 1995 WL 323710 (N.Y. Sup. Ct. 1995) (holding intermediary Prodigy was liable as a publisher for all posts made on its site, because it voluntarily deleted some forum postings).

²⁷ *Zeran v. America Online*, 129 F.3d 327 (4th Cir. 1997).

²⁸ *Id.* The quote continues: “Congress considered the weight of the speech interests implicated and chose to immunize service providers to avoid any such restrictive effect.”

²⁹ Eric Goldman, *The Ten Most Important Section 230 Rulings*, 20 TULANE J. TECH & I.P. ___ (2017), <https://ssrn.com/abstract=3025943>.



Yale Information Society Project

While there have long been worries about internet service providers favoring access to some content over others, there has been less concern about companies further along the pipeline holding an Internet on/off switch. In large part, this is because at other points in the pipeline, users have choice. Private companies can make their own rules, and consumers can choose among them. If GoDaddy won't register your domain, you can go to Bluehost or thousands of other companies. And while there may only be one Facebook, there are billions of other platforms and places online to post speech.

But the fewer choices you have for the infrastructure you need to stay online, the more serious the consequences when companies refuse or throttle service. This is one important reason net neutrality is so important. As Section 230 reveals, we all generally agree that it's appropriate for social media companies to take down certain kinds of content — that's how they ensure our newsfeeds aren't full of pornography or violence. But that doesn't mean we don't want that type of content to be able to exist *somewhere* on the Internet. Ensuring that ISPs remain content-neutral is necessary to guarantee that.³⁰

³⁰ This section borrows heavily from my article about the potential dangers in allowing internet infrastructure to regulate content. Kate Klonick, *The Terrifying Power of Internet Censors*, N.Y. TIMES (Sept. 13, 2017) <https://www.nytimes.com/2017/09/13/opinion/cloudflare-daily-stormer-charlottesville.html?>

Mr. LATTA. And, again, thank you for your testimony.
Ms. Moy, you are recognized for 5 minutes.

STATEMENT OF LAURA MOY

Ms. MOY. Thank you. Good morning, Chairmen Blackburn and Latta, Ranking Members Doyle and Schakowsky, and distinguished members of the subcommittees.

Consumers are frustrated. Ninety-one percent of adults feel that consumers have lost control of their personal information and nearly 70 percent think the law should do a better job of protecting their information. The law can do better, and it should do better. Consumers are in greatest need of greater control when they do not have a choice about whether to share the information in the first place. This is one reason that we have specific privacy laws that protect things like the information students share with educational institutions or the information patients share with doctors.

In these contexts and others, it is not permissible for companies to simply do what they wish with consumer information as long as they are transparent about it, something we see all too often online; rather, strong privacy protections apply by default. We need similar protection by default in other situations where information sharing is unavoidable, as well—for example, when consumer information is shared with a credit agency like Equifax or when consumer information is shared with the provider of an essential communication service like a broadband provider. We may also need protection by default for other types of online actors such as content platforms as they become bigger and more powerful and consumers increasingly find it unavoidable to share their information with those actors, as well. This is certainly a conversation worth having.

But whatever specific information-sharing problem or problems Congress decides to address, it should keep a few things in mind. First, Congress should not eliminate existing protections for consumers' information. This really should go without saying, but unfortunately, in an incredibly unpopular move earlier this year, Congress voted to eliminate strong Federal privacy rules that would have applied to broadband access providers.

Similarly, Congress has occasionally considered legislative proposals on data security and breach notification that would eliminate stronger State laws, but consumers want more protection for their information, not less. If Congress wishes to improve on the privacy and data security status quo, it should start by preserving the protections we already have. And just to touch for a second on net neutrality, the same applies in that context, as well.

Today's hearing is surfacing some concerns about the power platforms have to editorialize the things internet users read and say, but at the same time the FCC is considering wholesale elimination of rules that prevent broadband providers from doing that. Just imagine how much worse things could get if we start allowing broadband providers to muck with content. Again, consumers in this area need more protection, not less.

Second, prospective rulemaking authority is an incredibly important consumer protection tool. After-the-fact enforcement can be helpful, but an enforcement-only regime does not always create

clarity, and because it comes only after a problem has occurred, it does not necessarily protect consumers from the problem in the first place.

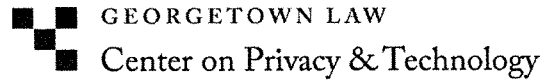
Granting rulemaking authority to an expert agency also fosters much-needed regulatory flexibility. We do not always know what the next privacy or data security threat will be, but unfortunately we all know that there will be one. An agency with rulemaking authority can respond to shifting threats more quickly than Congress.

Third, consumer protections are only as good as their enforcement, so any protections Congress creates on privacy or data security must be accompanied by strong enforcement authority. Right now, the FTC does substantial work on privacy and data security, but with few exceptions it does not have the ability to seek civil penalties for privacy and data security violations. In fact, FTC staff and commissioners have appeared before Congress requesting civil penalty authority to buttress their ability to enforce. Agencies that are tasked with protecting consumers' private information cannot do it without the proper tools. Civil penalty authority is needed.

Fourth, Congress should avoid the temptation to address complex challenges with a one-size-fits-all approach. There are different types of actors on the internet with different roles to play, different relationships with and commitments to consumers, different competition environments, and different abilities to solve problems. If we adopt a uniform regulatory approach to the entire internet, we are going to be left with the lowest common denominator, something like transparency with enforcement that just prohibits deceptive practices. That is not good enough. Consumers are asking for more.

I appreciate your commitment to this issue. I thank you, and I look forward to your questions.

[The prepared statement of Ms. Moy follows.]



**Statement of Laura Moy, Deputy Director
Center on Privacy & Technology at Georgetown Law**

Before the

**U.S. House of Representatives
Committee on Energy and Commerce
Subcommittee on Communications and Technology
and
Subcommittee on Digital Commerce and Consumer
Protection**

Hearing on

**Algorithms: How Companies' Decisions About Data and
Content Impact Consumers**

Wednesday, November 29, 2017

For more information, contact Laura Moy at laura.moy@georgetown.edu.

Introduction and Summary

Chairman Blackburn, Chairman Latta, Ranking Member Doyle, Ranking Member Schakowsky, and Members of the Subcommittees:

Consumers share information about themselves with others every day. In some instances, consumers have no choice but to share highly private information, such as when sharing is necessary to access an essential service. In other instances, consumers do have a choice, and share private information voluntarily. Private entities collect this consumer information because it is valuable, either on its own (such as in the case of a data broker intending to resell the information), or to power algorithmic decision-making. Algorithmic decision-making may streamline some aspects of our lives, but sometimes has flaws that lead to negative or unfair consequences.

Consumers feel that they have lost control of their private information, and consistently are asking for greater control. 91% of adults agree or strongly agree that consumers have lost control of how personal information is collected and used by companies, and 68% believe current laws are not good enough in protecting people's privacy online.

To foster the increased control over private information that consumers want, Congress should consider establishing protections that are forward-looking, flexible, strongly enforced, and appropriate based on context. In particular, agencies that are to be tasked with protecting consumers' private information must be given more powerful regulatory tools and stronger enforcement authority. But as Congress considers establishing new privacy and data security protections for consumers' private information, it should not eliminate existing protections.

Because we are still in the months following the massive Equifax breach, I also offer these Subcommittees a few targeted recommendations to better protect information held by credit reporting agencies (CRAs) First, Congress should enhance the authority of federal agencies to oversee the data security practices of consumer reporting agencies, to promulgate rules governing the data security obligations of financial institutions, and to enforce those obligations with civil penalties. Congress should also consider giving consumers better tools for redress when their personal information is compromised in a future breach by streamlining the credit freeze process, establishing protective tools for victims of child identity theft and medical identity theft, and prohibiting mandatory arbitration clauses.

I thank you for inviting me to testify on these important topics, and for your attention to privacy and data security.

1. Consumers share highly private information about themselves with a variety of actors on- and offline, and have varying degrees of choice with respect to that sharing

Consumers share information about themselves with others every day. In some instances, consumers have no choice but to share highly private information, for example to access an essential service. In other instances, consumers do have a choice, and share private information voluntarily.

A. Consumers have no choice but to share highly private information with an Internet service provider

Virtually every single consumer shares information about everything they do online with an Internet service provider (ISP). Consumers share this information not because they want to, but because they must. In the words of major ISP Comcast, “Internet service has become essential for success.”¹ Sharing information with an ISP is an unavoidable part of going online.

Making matters worse, many consumers cannot switch providers if they dislike the privacy practices of their ISP. In many areas, consumers have only one option when it comes to high-speed broadband. Even when there are two or three possible providers, switching costs—contract termination fees, installation fees, the time investment necessary to research and adopt an alternative—can make it very difficult for a subscriber of one provider to switch to another.

ISPs have tremendous visibility into nearly everything their clients do online, and can learn detailed information about consumers’ private lives. An ISP can see what websites its subscribers visit and when they visit them, and can make inferences based on that information. For example, domain names can expose details about health (plannedparenthood.org), finances (acecashexpress.com, particularly if accessed before each payday), political views (joinnra.nra.org), and other sensitive attributes.²

¹ Comcast, Internet Essentials Flyer, http://www.gaithersburgmd.gov/~media/city/documents/services/community/comcast_internet_essentials_flyer.pdf (last visited Apr. 6, 2017).

² The FCC’s Role in Protecting Online Privacy (Jan. 21, 2016) at 5, *available at* <https://www.newamerica.org/oti/policy-papers/the-fccs-role-in-protecting-online-privacy/>.

In addition, even when consumers' online activities have been purged of personal identifiers, such as name or a subscriber identifier, browsing histories can still be linked back to specific individuals. As explained by anonymization experts Sharad Goel and Arvind Narayanan, who recently presented a paper on the challenges of anonymizing web histories, "anonymous' web browsing records often contain an indelible mark of one's identity. We recruited nearly 400 users to send us their web browsing data stripped of any overt personal identifiers. In 70 percent of cases we could identify the individual from their web history alone."³

No other type of actor in the Internet ecosystem has access to as rich and reliable a stream of private information about individual users as ISPs. As noted privacy scholar Paul Ohm explained before the Senate Commerce Committee last year,

No other entity on the Internet possesses the same ability to see. If you are a habitual user of the Google search engine, Google can watch you while you search, and it can follow you on the first step you take away from the search engine. After that, it loses sight of you, unless you happen to visit other websites or use apps or services that share information with Google. If you are a habitual Amazon shopper, Amazon can watch you browse and purchase products, but it loses sight of you as soon as you shop with a competitor. Habitual Facebook users are watched by the company when they visit Facebook or use websites, apps or services that share information with Facebook, but they are not visible to Facebook at any other times.⁴

The threat to consumer privacy posed by ISPs is not something that consumers can address on their own. As I explained in an op-ed earlier this

³ Sharad Goel & Arvind Narayanan, *Why You Shouldn't Be Comforted by Internet Providers' Promises to Protect Your Privacy*, Future Tense (Apr. 4, 2017), http://www.slate.com/blogs/future_tense/2017/04/04/don_t_be_comforted_by_internet_providers_promises_to_protect_your_privacy.html (referring to Jessica Su, Ansh Shukla, Sharad Goel, & Arvind Narayanan, *Anonymizing Web Browsing Data with Social Networks*, available at <https://sharad.com/papers/twivacy.pdf>).

⁴ Testimony of Paul Ohm Before the Senate Commerce Committee, July 12, 2016, at 3, <http://paulohm.com/projects/testimony/PaulOhm20160712FCCPrivacyRulesSenate.pdf>.

year, none of the potential privacy protecting tools that consumers could use to hide their online activities from their ISP are perfect.⁵ Consumer-facing privacy options are weak, often difficult to locate, and even more difficult to understand. Tech-savvy consumers who can afford an additional monthly fee on top of what they already pay their ISP may consider signing up for a “virtual private network,” or VPN service, but that can be technically difficult for some consumers, as well as slow down the Internet experience. Consumers also can install a browser extension that will take the consumer to the encrypted version of a website whenever one is available, but many websites do *not* have encryption available, and even when encryption is available, it does not hide all private information from the ISP.

The bottom line when it comes to ISPs is that consumers have no choice but to share their information in order to get online.

B. Consumers have no choice but to share highly private information with credit reporting agencies

As with Internet service providers, consumers have no choice but to share highly private information with CRAs like Equifax. The massive troves of valuable and potentially damaging information that CRAs maintain are provided by furnishers, not by consumers themselves.

This is part of why consumers are so outraged by the recent Equifax breach. The 165.5 million Americans whose private details were breached in the Equifax attack now face an increased risk of identity theft in perpetuity. Now that their names, Social Security numbers, and other difficult-to-change data closely tied to financial records have been breached, those details are out there forever—there is no putting the genie back in the bottle.

And there is no question that, entrusted with this private information through no affirmative choice by consumers, Equifax made serious mistakes. Equifax could and should have prevented a breach of this magnitude from occurring. Indeed, the scale of the breach alone—affecting some 45% of American consumers in an attack that took place over the course of months—indicates that Equifax’s security program was riddled with problems. And it was. Equifax’s unreasonable security failures include the failure to encrypt

⁵ Laura Moy, *Think You Can Protect Your Privacy from Internet Providers Without FCC Rules? Good Luck.*, The Daily Dot (Mar. 28, 2017), <https://www.dailydot.com/layer8/congress-kill-isp-privacy-protections/>.

the large volume of data that ultimately was exfiltrated by attackers,⁶ the months-long failure to patch the critical Apache Struts vulnerability that was exploited,⁷ the apparent lack of appropriate management and redundancies to ensure the patch would be applied,⁸ and the months-long failure to detect the breach even as attackers continued to access and steal sensitive consumer data.

Even though many consumers may have lost or diminished trust in Equifax—and perhaps other CRAs as well—following the Equifax breach, the decision to share private information with CRAs is out of consumers' hands.

C. Consumers often do have a choice whether or not to share private information

Although in some instances, such as where ISPs or CRAs are concerned, consumers have no choice but to share private information, consumers also are often asked or invited to share information about themselves in circumstances where such sharing would be completely voluntary. For example, consumers sometimes—but not always—are willing to participate in voluntary surveys in which they are asked to share information about their preferences or habits. Consumers also may share information with an online discussion forum so that they can participate in forum conversations, or with a shopping list application so that they can keep better track of groceries they need to purchase.

⁶ *Oversight of the Equifax Data Breach: Answers for Consumers: Hearing Before the H. Comm. on Energy and Commerce Subcomm. on Digital Commerce and Consumer Protection*, 115th Cong. (Oct. 3, 2017) (statement of Richard F. Smith, Former Chairman and CEO, Equifax, Inc.), preliminary transcript at 81, available at <http://docs.house.gov/meetings/IF/IF17/20171003/106455/HHRG-115-IF17-Transcript-20171003.pdf> (“To be very specific this data was not encrypted at rest.”)[hereinafter *Oct. 3 Hearing*]

⁷ See Lily Hay Newman, *Equifax Officially Has No Excuse*, WIRED (Sept. 14, 2017), <https://www.wired.com/story/equifax-breach-no-excuse/>.

⁸ *Oct. 3 Hearing* (statement of Richard F. Smith, Former Chairman and CEO, Equifax, Inc.), preliminary transcript at 35, (“The human error was the individual who is responsible for communicating in the organization to apply the patch did not.”); see Russell Brandom, *Former Equifax CEO Blames Breach on a Single Person Who Failed to Deploy Patch*, The Verge (Oct. 3, 2017), <https://www.theverge.com/2017/10/3/16410806/equifax-ceo-blame-breach-patch-congress-testimony>.

2. Information collected from and about consumers is used to power algorithmic decision-making that can be problematic

Information about consumers is not collected in a vacuum; private entities collect consumer information because it is valuable, either on its own (such as in the case of a data broker intending to resell the information), or to power automated decision-making. Indeed, many things that once were decided by humans are now often decided—or at least influenced—by predictive formulas designed by data scientists, and those formulas may be responsible for decisions that have important effects on consumers' lives. Algorithms may be used to determine which job applicants are invited to come in for an interview,⁹ where police officers should patrol,¹⁰ or how long a person convicted of a crime should spend in jail.¹¹ Algorithms also select much of what we read and see online. They may determine which products are presented to us in advertisements, which movies are recommended to us, which friends' photos we see, and which news articles we read.

Algorithmic decision-making may streamline some aspects of our lives, but algorithms can sometimes have flaws that lead to negative or unfair consequences. For example, hiring algorithms have been accused of unfairly discriminating against people with mental illness.¹² Sentencing algorithms—intended to make sentencing fairer by diminishing the role of potentially biased human judges—may actually discriminate against Black people.¹³ Search algorithms may be more likely to surface advertisements for arrest

⁹ Lauren Weber & Elizabeth Dwoskin, *Are Workplace Personality Tests Fair? Growing Use of Tests Sparks Scrutiny Amid Questions of Effectiveness and Workplace Discrimination*, W.S.J. (Sept. 29, 2014), <https://www.wsj.com/articles/are-workplace-personality-tests-fair-1412044257>.

¹⁰ Laurel Eckhouse, *Big Data May Be Reinforcing Racial Bias in the Criminal Justice System*, Wash. Post (Feb. 10, 2017), https://www.washingtonpost.com/opinions/big-data-may-be-reinforcing-racial-bias-in-the-criminal-justice-system/2017/02/10/d63de518-ee3a-11e6-9973-c5efb7ccfb0d_story.html.

¹¹ Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, *Machine Bias*, ProPublica (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

¹² Weber & Dwoskin, *supra* note 9.

¹³ Angwin, et al., *supra* note 11.

records—regardless of whether such records exist—when presented with characteristically Black names.¹⁴

The use of consumer data to power algorithmic decision-making deserves particularly close scrutiny when the decisions to be made will affect opportunities for education, healthcare, financial products, or employment. For example, policymakers may reasonably not be concerned with flawed algorithms that display ads for wine to the wrong crowd, but there is greater cause for concern when a study shows—as one has—that male job seekers are much more likely than equivalent female jobs seekers to be shown ads for high-paying executive ads.¹⁵

It may also be problematic when consumer data is used to power the targeted distribution of content that may distort consumers' perception of issues of importance, such as political issues. This is especially the case when consumers are not aware that algorithms are at work personalizing which content they will see and in what order.¹⁶ Consider, for example, a hypothetical posed by digital analytics consultant Angela Grammatas:

[I]magine that “Jane Internet” loves cats, and visits cats.com daily. One day, she’s considering how to vote on a local proposition, and she does some research by visiting two political news sites at opposite ends of the spectrum. She reads a relevant article on each site, getting a balanced view of the issue. Let’s imagine that the “Yes on Prop A” campaign has access to

¹⁴ Latanya Sweeney, *Discrimination in Online Ad Delivery*, Communications of the Association of Computing Machinery (Jan. 2013).

¹⁵ Tom Simonite, *Probing the Dark Side of Google’s Ad-Targeting System*, MIT Technology Review (July 6, 2015), <https://www.technologyreview.com/s/539021/probing-the-dark-side-of-googles-ad-targeting-system/>.

¹⁶ One study of 40 Facebook users found that a majority of participants—62.5%—did not know that content on Facebook was filtered. According to the study’s authors, “In [the unaware users’] opinion, missing a public story was due to their own actions, rather than those of Facebook. Importantly, these participants felt that they missed friends’ stories because they were scrolling too quickly or visiting Facebook too infrequently.” Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, & Christian Sandvig, *“I Always Assumed that I Wasn’t Really that Close to [Her]”: Reasoning About Invisible Algorithms in the News Feed*, in CHI ’15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems at 153, 156 (New York 2015).

retargeting capabilities that utilize that large, blended dataset. Soon, Jane starts to see “Vote Yes on Prop A” advertisements on many unrelated websites, with the message that Prop A will be great for local wildlife.

Jane has no way of knowing this, but that pro-wildlife message has been chosen specifically for her, because of her past visits to cats.com. The ads are everywhere online (for Jane), so Jane believes that this message is a primary “Yes on A” talking point, and she’s encouraged to vote in agreement. The “No on A” campaign never has any opportunity to discuss or debate the point. They may not even know that the cats-related topic has been raised, because they’ve never even been exposed to it—that message is reserved for retargeting campaigns directed at people like Jane. Jane’s attempt to be a well-informed voter has been usurped by retargeting. And, perhaps most importantly, Jane doesn’t even know this has happened.¹⁷

Even when the use of consumer data to power algorithmic decision-making can be directly harmful, such as when it affects livelihood-related opportunities or distorts consumers’ perception of issues of importance, it may still be considered privacy violative when it exceeds consumers’ expectations about how the data would be used.

3. Protections for consumers’ private information should be forward-looking, flexible, strongly enforced, and carefully tailored based on context

Consumers want more control over their private information, and consistently are asking for it. According to a 2016 report from the Pew Research Center, “91% of adults agree or strongly agree that consumers have lost control of how personal information is collected and used by companies,” and 68% believe current laws are not good enough in protecting people’s

¹⁷ Angela Grammatas, *Guest Post: Make Your Browsing Noiszy*, Mathbabe (Mar. 31, 2017), <https://mathbabe.org/2017/03/31/guest-post-make-your-browsing-noiszy/>.

privacy online.¹⁸ Consumers need clear forward-looking protections that are flexible, strongly enforced, and appropriate based on context.

A. Protections for consumers' private information should be forward-looking, flexible, and strongly enforced

The FTC brings the bulk of federal privacy enforcement actions, but it lacks the tools it needs to be as effective as it could be. The agency only has after-the-fact enforcement authority, but no ability to define rules of the road before consumer data is used in ways that consumers consider inappropriate. And apart from the few contexts in which it has specific privacy authority, the FTC generally can only take enforcement action against entities that use consumer information in ways that violate their own consumer-facing commitments. When the FTC does take action to enforce, it is generally unable to pursue penalties that would serve as an effective punishment for violators, and an effective deterrent for others.¹⁹ To improve privacy and data security for consumers, the FTC—or another agency or agencies—must be given more powerful regulatory tools and stronger enforcement authority.

The law should grant an expert agency or agencies the authority to develop prospective privacy and data security rules, in consultation with the public, so that data collectors and users can know in advance what standards apply to consumers' information.

Regulations should also be flexible, allowing agencies to adjust them as technology changes, as the FTC did just a few years ago with the COPPA Rule.²⁰ Consumers are constantly encountering new types of privacy and data

¹⁸ Lee Rainie, Pew Research Center, *The State of Privacy in Post-Snowden America* (Sept. 21, 2016), <http://www.pewresearch.org/fact-tank/2016/09/21/the-state-of-privacy-in-america/>.

¹⁹ There are exceptions to this rule. As the FTC explains, “If a company violates an FTC order, the FTC can seek civil monetary penalties for the violations. The FTC can also obtain civil monetary penalties for violations of certain privacy statutes and rules, including the Children’s Online Privacy Protection Act, the Fair Credit Reporting Act, and the Telemarketing Sales Rule.” FTC, *Privacy & Security Update 2016*, <https://www.ftc.gov/reports/privacy-data-security-update-2016>.

²⁰ Federal Trade Commission, *FTC Strengthens Kids’ Privacy, Gives Parents Greater Control over Their Information by Amending Children’s Online Privacy Protection Rule* (Dec. 19, 2012), <https://www.ftc.gov/news-events/>

security threats as the information landscape evolves. Where flexibility exists, policymakers use it to respond to changing threats. For example, states adjust data security and breach notification protections as changing circumstances require, such as by extending protection to additional categories of information, including medical information and biometric data.²¹ We can't always forecast the next big threat years in advance, but unfortunately, we know that there will be one.

Congress also should ensure that whatever agency or agencies are to be in charge of enforcing privacy and data security standards have substantial civil penalty enforcement authority. Indeed, the FTC has repeatedly asked for the civil penalty authority it needs to enforce data security.²² Regulations are effective to deter violations only if entities fear the punishment that would surely follow.

B. Protections for consumers' private information should be tailored based on the avoidability of the information sharing,

press-releases/2012/12/ftc-strengthens-kids-privacy-gives-parents-greater-control-over.

²¹ William Elser, *Recent Updates to State Data Breach Notification Laws in New Mexico, Tennessee, Virginia*, Lexology (May 1, 2017), <https://www.lexology.com/library/detail.aspx?g=b02a15ac-a3c3-460d-bc5e-1d29778c4e59> (“New Mexico’s new law defines ‘personal identifiable information’ consistently with most other states, and joins a growing number of states that have broadened the definition to include ‘biometric data,’ which is defined to include ‘fingerprints, voice print, iris or retina patterns, facial characteristics or hand geometry.’”).

²² See, e.g., Testimony of Jessica Rich, Federal Trade Commission, before the House Oversight and Government Reform Committee Subcommittees on Information Technology and Health, Benefits, and Administrative Rules regarding Opportunities and Challenges in Advancing Health Information Technology (Mar. 22, 2016) at 7, *available at* <https://oversight.house.gov/wp-content/uploads/2016/03/2016-03-22-Rich-Testimony-FTC.pdf>; Maureen Ohlhausen, Commissioner, Fed. Trade Comm’n, Remarks Before the Congressional Bipartisan Privacy Caucus (Feb. 3, 2014), transcript *available at* https://www.ftc.gov/system/files/documents/public_statements/remarks-commissioner-maureen-k.ohlhausen/140203datasecurityohlhausen.pdf.

the sensitivity of the information, and the expectations of consumers

There is no one-size-fits-all approach for privacy. Rather, privacy laws and regulations should be context-specific, carefully tailored based on the avoidability of the information sharing, the sensitivity of the information shared, and the expectations of consumers.

Whether a consumer has the ability to avoid sharing personal information with a private entity—such as in the case of a shopping list application, or no choice—such as in the case of an ISP or CRA, is relevant in considering what level of privacy protection is appropriate for a particular context. When information sharing is unavoidable or less avoidable by consumers, it is important that the information be protected. This explains in part why there are a variety of laws that protect consumer information in specific contexts in which sharing is unavoidable—such as the information shared by students in an educational context,²³ by consumers in a financial context,²⁴ by customers in a telecommunications context,²⁵ and by patients in a medical context.²⁶

This is also consistent with the FTC's evaluation of potentially problematic data-related practices under its Section 5 authority to prohibit unfair practices. When considering whether a practice is unfair, the FTC asks not only whether the practice is harmful, but also whether the practice is one that consumers can avoid. In its policy statement on unfairness, the FTC explained,

Normally we expect the marketplace to be self-correcting, and we rely on consumer choice—the ability of individual consumers to make their own private purchasing decisions without regulatory intervention—to govern the market. We anticipate that consumers will survey the available alternatives, choose those that are most desirable, and avoid those that are inadequate or unsatisfactory. However, it has long been recognized that certain types of sales techniques may prevent

²³ Family Educational Rights and Privacy Act, 20 U.S.C. § 1232g.

²⁴ Gramm-Leach-Bliley Act, Pub. L. No. 106-102, 113 Stat. 1338, (1999).

²⁵ 47 U.S.C. § 222.

²⁶ Health Insurance Portability and Accountability Act of 1996, Pub. L. 104-191, 110 Stat. 1936 (1996).

consumers from effectively making their own decisions, and that corrective action may then become necessary. Most of the Commission's unfairness matters are brought under these circumstances. They are brought, not to second-guess the wisdom of particular consumer decisions, but rather to halt some form of seller behavior that unreasonably creates or takes advantage of an obstacle to the free exercise of consumer decisionmaking.²⁷

Whether or not information sharing is avoidable by a consumer is often tied to the question of whether or not a service or transaction is essential. When a service is essential—such as with Internet connectivity—information sharing may be considered unavoidable because the consumer cannot reasonably decline the service altogether.

Policymakers should also consider how the avoidability of any particular choice presented to a consumer may be affected or distorted by other factors that make it unavoidable as a practical matter, such as whether the choice is technically difficult for most consumers to understand or exercise, whether network effects diminish consumers' perception of the choice as optional, whether well-documented cognitive biases inhibit consumers' ability to rationally evaluate potential risks associated with the

²⁷ FTC, *FTC Policy Statement on Unfairness* (Dec. 17, 1980), <https://www.ftc.gov/public-statements/1980/12/ftc-policy-statement-unfairness>.

choice,²⁸ or whether the entity collecting consumer information is using coercive or deceptive tactics to get consumers to exercise a particular choice.²⁹

In determining what level of protection should be afforded to information shared in a particular context, policymakers should also examine how sensitive the shared information is. For example, the Children’s Online Privacy Protection Act recognizes that information about children deserves heightened protection.³⁰ Other laws recognize the heightened sensitivity of health information³¹ and financial information.³² In the past, the question of sensitivity has often been the most important in considering how well the law should protect consumers’ information. Data analysis techniques have advanced over time, however, and it is becoming clear that classically sensitive information can often be deduced from categories of information not traditionally thought of as sensitive. For example, as computer scientist Ed Felten explained in testimony before the Senate Judiciary Committee regarding telephone metadata, “Calling patterns can reveal when we are awake and asleep; our religion . . . our work habits and our social attitudes; the number of friends we have; and even our civil and political affiliations.”³³

²⁸ See Alessandro Acquisti, *Privacy in Electronic Commerce and the Economics of Immediate Gratification*, in EC ’04 Proceedings of the 5th ACM Conference on Electronic Commerce (New York 2004), at 21, 27 (“We have shown that a model of rational privacy behavior is unrealistic, while models based on psychological distortions offer a more accurate depiction of the decision process. We have shown why individuals who genuinely would like to protect their privacy may not do so because of psychological distortions well documented in the behavioral economics literature. We have highlighted that these distortions may affect not only naïve individuals but also sophisticated ones. Surprisingly, we have also found that these inconsistencies may occur when individuals perceive the risks from not protecting their privacy as significant.”).

²⁹ See Lauren E. Willis, *When Nudges Fail: Slippery Defaults*, 80 U. Chi. L. Rev. 1155 (2012); Lauren E. Willis, *Why Not Privacy by Default?*, 29 Berkeley Tech. L.J. (2014).

³⁰ 15 U.S.C. §§ 6501–6506.

³¹ *E.g.* Health Insurance Portability and Accountability Act of 1996, Pub. L. 104–191, 110 Stat. 1936 (1996).

³² *E.g.* Gramm-Leach-Bliley Act, Pub. L. No. 106–102, 113 Stat. 1338, (1999).

³³ *Continued Oversight of the Foreign Intelligence Surveillance Act: Hearing before the S. Comm. on the Judiciary*, 113th Cong. 8–10 (2013) (statement of Edward Felten, Professor of Computer Science and Public Affairs, Princeton

Last year the FTC found that television viewing history can be considered sensitive information,³⁴ and the Federal Communications Commission (FCC) found that web browsing history can be considered sensitive.³⁵ Indeed, patent applications filed by Google indicate that it is possible to estimate user demographics and location information based on browsing histories.³⁶

Protection for consumers' information should also be tailored based on consumers' expectations for how the information will be used.

C. Congress should not eliminate existing protections for consumers' information

As Congress considers establishing new privacy and data security protections for consumers' private information, it should not eliminate existing protections. Americans are asking for *more* protections for their private information, not less. This explains why when this body voted earlier this year to eliminate strong privacy regulations that had recently been passed by the FCC, consumers—on both sides of the aisle—were outraged.³⁷

University) *available at* <http://www.judiciary.senate.gov/meetings/continued-oversight-of-the-foreign-intelligence-surveillance-act>.

³⁴ Complaint at ¶ 32, *FTC v. Vizio*, Case No. 2:17-cv-00758, D.N.J. (filed Feb. 6, 2017), *available at* https://www.ftc.gov/system/files/documents/cases/170206_vizio_2017.02.06_complaint.pdf.

³⁵ Federal Communications Commission, *Fact Sheet: The FCC Adopts Order to Give Broadband Consumers Increased Choice over Their Personal Information*, https://apps.fcc.gov/edocs_public/attachmatch/DOC-341938A1.pdf.

³⁶ See U.S. Patent Application No. 13/652,198, Publication No. 20130138506 (published May 30, 2013)(Google Inc., applicant)(“demographics data may include a user's age, gender, race, ethnicity, employment status, education level, income, mobility, familial status (e.g., married, single and never married, single and divorced, etc.), household size, hobbies, interests, location, religion, political leanings, or any other characteristic describing a user or a user's beliefs or interests.”); U.S. Patent Application No. 14/316,569, Publication No. 20140310268 (published Oct. 16, 2014)(Google Inc., applicant).

³⁷ See Matthew Yglesias, *Republicans' Rollback of Broadband Privacy Is Hideously Unpopular*, *Vox* (Apr. 4, 2017), <https://www.vox.com/policy-and-politics/2017/4/4/15167544/broadband-privacy-poll>.

Some lawmakers argued that repeal of the FCC's rules was needed to foster development of a consistent approach to privacy across the Internet.³⁸ But as FTC Commissioner Terrell McSweeney noted, "If consistency were truly the goal, then we would likely increase protections for privacy, rather than unraveling them. That is the policy conversation we ought to be having—instead we are fighting a rear-guard action defending basic protections."³⁹

Congress also should not eliminate existing consumer protections at the state level. State laws play an important role in filling gaps that exist in federal legislation, and state attorneys general play an important role in enforcing privacy and data security standards. For example, in data security and breach notification, some state laws protect categories of information that are not protected by other states, and would not be protected by a number of proposals for federal data security and breach notification legislation.⁴⁰ State attorneys general play a critical role in policing data security and guiding breach notification to match the needs of their own residents, and are essential in conducting ongoing monitoring after a breach has occurred to help protect residents from any aftermath, especially where small data breaches are concerned. According to the Massachusetts State Attorney General's Office, Massachusetts alone saw 2,314 data breaches

³⁸ See Alex Byers, *House Votes to Revoke Broadband Privacy Rules*, Politico (Mar. 28, 2017), <https://www.politico.com/story/2017/03/house-votes-to-revoke-broadband-privacy-rules-236607>.

³⁹ Terrell McSweeney, Commissioner, Fed. Trade Comm'n, Remarks on "*The Future of Broadband Privacy and the Open Internet: Who Will Protect Consumers?*" (Apr. 17, 2014), at 4, https://www.ftc.gov/system/files/documents/public_statements/1210663/mcsweeney_-_new_americas_open_technology_institute_4-17-17.pdf.

⁴⁰ See Testimony of Laura Moy before the House Energy & Commerce Committee Subcommittee on Commerce, Manufacturing, and Trade regarding the Data Security and Breach Notification Act of 2015 (Mar. 11, 2015) at 3–5, *available at* <https://democrats-energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/documents/Testimony-Moy-CMT-Data-Breach-Legislation-2015-03-18.pdf>; see also Responses to Additional Questions for the Record of Laura Moy before the House Energy & Commerce Committee Subcommittee on Commerce, Manufacturing, and Trade, <http://docs.house.gov/meetings/IF/IF17/20150318/103175/HHRG-114-IF17-Wstate-MoyL-20150318.pdf>.

reported in 2013, 97% of which involved fewer than 10,000 affected individuals.⁴¹ Each data breach affected, on average, 74 individuals.⁴²

4. Specific recommendations for regulation of CRAs

Congress should advance federal legislation to subject CRAs to closer regulatory oversight and stronger enforcement, and to enhance consumers' control of their own personal information.

A. Congress should consider subjecting the security practices of consumer reporting agencies to closer regulatory oversight and stronger enforcement

First and foremost, Congress should consider vesting a federal agency or agencies with the authority to more closely regulate and enforce the data security practices of CRAs. Both the FTC and the Consumer Financial Protection Bureau announced they were looking into the Equifax breach shortly after it occurred. But to help prevent similar breaches from occurring in the future, Congress should explore bolstering these agencies' authority to promulgate rules governing the data security practices of CRAs, to conduct ongoing review of CRAs' data security practices, to enforce rules, and to seek civil penalties for violations.

At this point, the FTC has rulemaking and enforcement authority over CRAs' data security practices, but no supervisory authority. In accordance with the Gramm-Leach-Bliley Act (GLBA), in 2002 the FTC promulgated the Safeguards Rule,⁴³ which governs the data security obligations of financial institutions, including CRAs.⁴⁴ Companies covered by the rule not only must align their own data security practices with the requirements of the rule, but

⁴¹ Testimony of Sara Cable before the House Energy & Commerce Committee Subcommittee on Commerce, Manufacturing, and Trade regarding the Data Security and Breach Notification Act of 2015, *available at* <http://docs.house.gov/meetings/IF/IF17/20150318/103175/HHRG-114-IF17-Wstate-CableS-20150318.pdf>.

⁴² *Id.*

⁴³ 16 C.F.R. §314.

⁴⁴ Fed. Trade Comm'n, *Financial Institutions and Customer Information: Complying with the Safeguards Rule*, <https://www.ftc.gov/tips-advice/business-center/guidance/financial-institutions-customer-information-complying> (last visited Oct. 23, 2017).

also must ensure that their affiliates and service providers safeguard customer information in their care.⁴⁵ But as the Congressional Research Service explains, the FTC “has little up-front supervisory or enforcement authority, making it difficult to prevent an incident from occurring and instead often relying on enforcement after the fact.”⁴⁶

The CFPB, on the other hand, has exercised supervisory authority over CRAs since 2012, but lacks the authority to promulgate rules implementing or to enforce the data security provisions of GLBA.⁴⁷ Title X of the Dodd-Frank Act granted the CFPB rulemaking authority for much of GLBA, but according to the CFPB itself, Dodd-Frank “excluded financial institutions’ information security safeguards under GLBA Section 501(b) from the CFPB’s rulemaking, examination, and enforcement authority.”⁴⁸

In addition, Congress should consider urging the FTC and/or CFPB to complete a notice and comment rulemaking process to update the Safeguards Rule. The existing Safeguards Rule was promulgated in 2002. In 2016 the FTC began the process of updating that rule, and solicited public comment on a number of both questions, including about the substantive standards set forth in the rule, such as, “Should the Rule be modified to include more specific and prescriptive requirements for information security plans?” and “Should the Rule be modified to reference or incorporate any other information security standards or frameworks, such as the National Institute of Standards and Technology’s Cybersecurity Framework or the Payment Card Industry Data Security Standards?”⁴⁹ The FTC has not completed the update. Most recently, in June, the FTC published a notice indicating that

⁴⁵ *Id.*

⁴⁶ N. Eric Weiss, *The Equifax Data Breach: An Overview and Issues for Congress*, CRS Insight (Sept. 29, 2017) at 2.

⁴⁷ *Id.*

⁴⁸ Consumer Fin. Protection Bureau, *Privacy of Consumer Financial Information – Gramm-Leach-Bliley Act (GLBA) Examination Procedures* at 1 (Oct. 2016), https://s3.amazonaws.com/files.consumerfinance.gov/f/documents/102016_cfpb_GLBAExamManualUpdate.pdf.

⁴⁹ FTC Standards for Safeguarding Customer Information, Request for Public Comment, 81 Fed. Reg. 173 (Sept. 7, 2016), https://www.ftc.gov/system/files/documents/federal_register_notices/2016/09/frn_standards_for_safeguarding_customer_informtion.pdf.

the Safeguards Rule is “currently under review,” and that the agency does not expect to complete the review in 2017.⁵⁰

Congress should also consider giving one or both agencies the authority to seek civil penalties for violations of the Safeguards Rule. The FTC has itself called for civil penalty authority in the past to buttress its data security authority. As now-Acting Chairman of the FTC (then a Commissioner) Maureen Ohlhausen argued in remarks she delivered before Congressional Bipartisan Privacy Caucus in 2014,

Legislation in both areas—data security and breach notification—should give the FTC the ability to seek civil penalties to help deter unlawful conduct, rulemaking authority under the Administrative Procedure Act, and jurisdiction over non-profits. Under current laws, the FTC only has the authority to seek civil penalties for data security violations with regard to children’s online information under COPPA or credit report information under the FCRA.⁵¹ To help ensure effective deterrence, we urge Congress to allow the FTC to seek civil penalties for data security and breach notice violations in appropriate circumstances.⁵²

To improve the FTC’s and CFPB’s ability to protect Americans from poor data security practices of financial institutions that house extremely sensitive information, Congress should consider vesting one or both agencies with full-throated supervisory, rulemaking, and enforcement authority, and consider urging the update of the Safeguards Rule.

⁵⁰ FTC Regulatory Review Schedule, 82 Fed. Reg. 123 (June 28, 2017), https://www.ftc.gov/system/files/documents/federal_register_notices/2017/06/reg_review_schedule_published_frn.pdf.

⁵¹ The FTC can also seek civil penalties for violations of administrative orders. 15 U.S.C. § 45(d) (footnote in original).

⁵² Maureen Ohlhausen, Commissioner, Fed. Trade Comm’n, Remarks Before the Congressional Bipartisan Privacy Caucus (Feb. 3, 2014), transcript *available at* https://www.ftc.gov/system/files/documents/public_statements/remarks-commissioner-maureen-k.ohlhausen/140203datasecurityohlhausen.pdf.

B. Congress should consider expanding consumer tools for redress in the event of a CRA breach

In addition to taking steps to bolster regulatory and enforcement authority to help prevent similar breaches from taking place in the future, Congress should consider giving consumers better tools for redress when their personal information is compromised in a future breach. Specifically, Congress should consider streamlining the credit freeze process, establishing protective tools for victims of child identity theft and medical identity theft, and prohibiting mandatory arbitration clauses.

The credit freeze process is overdue for an overhaul—although credit freezes offer useful protection, they can be tedious, inconvenient, and costly. The credit freeze is, according to U.S. PIRG, “your best protection against someone opening new credit accounts in your name,”⁵³ and the IRS encourages consumers to consider requesting a freeze “if you were part of a large-scale data breach.”⁵⁴ But the FTC cautions consumers considering a credit freeze to “[c]onsider the cost and hassle factor,” because a credit freeze can delay access to credit, is only truly effective if secured across all three major CRAs, and may come at a cost of \$5 to \$10 for each CRA every time a consumer wishes to freeze or thaw their credit.⁵⁵ Congress should consider requiring CRAs to make it faster, easier, and free for consumers to freeze or thaw their credit, and to work together to ensure that a credit freeze or thaw request made with one CRA is applied to other bureaus as well. A protective tool like the credit freeze should be simplified so that consumers can easily access it, and should not be made available only to those consumers who can afford to pay for it either in time or in dollars.

Congress should also consider expanding the suite of tools that the law requires be made available to help consumers who become victims of identity

⁵³ Mike Litt & Edmund Mierzwinski, U.S. PIRG, *Why You Should Get Credit Freezes Before Your Information Is Stolen: Tips to Protect Yourself Against Identity Theft & Financial Fraud* at 1 (Oct. 2015), available at https://uspirg.org/sites/pirg/files/reports/USPIRGFREEZE_0.pdf.

⁵⁴ Internal Revenue Service, *Tips for Using Credit Bureaus to Help Protect Your Financial Accounts*, <https://www.irs.gov/newsroom/tips-for-using-credit-bureaus-to-help-protect-your-financial-accounts> (last visited Oct. 23, 2017).

⁵⁵ Lisa Weintraub Schifferle, Fed. Trade Comm’n, *Fraud Alert or Credit Freeze – Which Is Right for You?* (Sept. 14, 2017), <https://www.consumer.ftc.gov/blog/2017/09/fraud-alert-or-credit-freeze-which-right-you> (last visited Oct. 23, 2017).

theft. For consumers of financial identity theft, there are modest protections in place, including enhanced free credit monitoring and fraud alert options. But for other forms of identity theft, such as child identity theft and medical identity theft, no such tools exist. Congress should consider providing these victims with the tools they'll need to protect their identity—and if stolen, restore it.

In addition, Congress should consider prohibiting the use of mandatory arbitration clauses designed to keep consumers who have been the victim of data security or privacy violations out of court. Equifax invited tremendous criticism for its inclusion of a forced arbitration clause in the terms made available to individuals subject to its breach, and has since stated that it never intended to include the arbitration clause.⁵⁶ Congress should make clear that mandatory arbitration is never permissible where the privacy and data security obligations of financial institutions are concerned.

5. Congress should not issue federal data security or breach notification legislation that eliminates existing consumer protections

If Congress considers passing federal legislation on data security and breach notification, consumers would best be served by a bill that does not preempt state laws. If Congress nevertheless considers legislation that does preempt state data security and breach notification provisions, I urge you to explore legislation that is narrow, and that merely sets a floor for disparate state laws—not a ceiling.

In the event, however, that Congress nevertheless seriously considers broadly preemptive data security and breach notification legislation, the new federal standard should strengthen, or at the very least preserve, important protections that consumers currently enjoy at the state level. In particular, federal legislation:

- 1) should not ignore the serious physical, emotional, and other non-financial harms that consumers could suffer as a result of misuses of their personal information,

⁵⁶ *Oct. 3 Hearing* (prepared testimony of Richard F. Smith, Former Chairman and CEO, Equifax, Inc.), at 5, <http://docs.house.gov/meetings/IF/IF17/20171003/106455/HHRG-115-IF17-Wstate-SmithR-20171003.pdf>.

- 2) should not eliminate data security and breach notification protections for types of data that are currently protected under state law,
- 3) should provide a means to expand the range of information protected by the law as technology develops,
- 4) should include enforcement authority for state attorneys general, and
- 5) should be crafted in such a way as to avoid preempting privacy and general consumer protection laws.

I have previously presented these arguments before this Committee,⁵⁷ so I will not elaborate on them here.

6. Conclusion

I am grateful for the Subcommittees' attention to these important issues, and for the opportunity to present this testimony.

⁵⁷ Testimony of Laura Moy before the House Energy & Commerce Committee Subcommittee on Commerce, Manufacturing, and Trade regarding the Data Security and Breach Notification Act of 2015 (Mar. 11, 2015), *available at* <https://democrats-energycommerce.house.gov/sites/democrats-energycommerce.house.gov/files/documents/Testimony-Moy-CMT-Data-Breach-Legislation-2015-03-18.pdf>.

Mr. LATTI. And again, thank you for your testimony this morning.

And Dr. Tucker, you are recognized for 5 minutes.

STATEMENT OF CATHERINE TUCKER

Dr. TUCKER. So, first of all, I would just like to say what a huge honor it is to be invited here today. Thank you very much for the invitation. What I want to do in my 5 minutes is, first of all, talk about some research I did into an apparent algorithmic bias and then talk about three implications for policy.

Now, this particular research topic—what we did was we ran a field test on Facebook where we placed an ad which advertised job opportunities in science and technology. And we placed that ad, we also replicated it on Google and Twitter, and we found that the advertising algorithm ended up showing this ad for job opportunities in science to 40 percent more men than women. And on the face of it, this seems really quite concerning because obviously this is an area where we would like parity of gender opportunity.

Now, I say on the face of it, it sounds concerning, because our research didn't stop there, which is usually how research stops, but instead we actually delved into the reasons why this apparent discrimination had happened. And we ruled out the usual leading explanations, which is either that humans are biased, absorb cultural prejudice, or the idea that somehow women have self-inflicted not seeing the ad on themselves by not reacting to it. Instead, if women ever saw the ad, they loved it. They clicked on it.

Instead, what actually was going on is all in terms of understanding how the algorithm works, which is that an advertising algorithm basically runs an auction in real time where advertisers bid for eyeballs, and there were some advertisers out there that liked to show ads just to women, and as a result they pay more to show the ad to women. And because we had set up our ad to be gender-neutral, the algorithm thought it was doing us a favor by trying to minimize our costs and not show our ad to those expensive female eyeballs, but instead prioritize those cheaper male ones.

Now, that takes us, you know, to show that actually economic forces actually shape a lot, you know, how we see algorithms work. And I want to just highlight three implications of policy. The first implication is that about algorithmic transparency. Now, algorithmic transparency just sounds wonderful, right? Who could ever argue with transparency?

But, in this case, let's suppose we could ever decode the pages and pages of algorithms which underlie this ad auction. All we would find is an innocent algorithm trying to save advertisers money. It wouldn't give us really any insight into the potential for bias, and I think that is another argument to build on what we have heard earlier, why transparency, though just so beautifully sounding, is probably not a solution here.

The second thing I want to emphasize is, it may be tempting, and we sort of, you know, we have heard a little bit of this idea that maybe the problem is not the algorithms, it is the data that feeds them. And I do want to caution the committee surrounding just simply restricting data flows in this economy. I have done

some research. I have testified it into the past about the really quite hideous effects that attempts to regulate privacy in online advertising have had on the health and strength of the technology industry in Europe.

We show that they had a 66 percent drop in efficiency after passing regulation, and you just have to sort of fast forward 10 years, look at the strength of the American tech industry relative to Europe to see where that has led. I have also done some research in the U.S. We should emphasize that just restricting data in the health arena has actually led to some really quite negative consequences, such as hospitals failing to adopt potentially lifesaving neonatal technology saving babies.

Now, the last—so that is why I am worried about restricting data as a solution—the last thing I just want to say is, look, in some sense you could write a headline saying “MIT professor finds ad algorithm doesn’t show job ads to women,” but imagine if I had found that for toothpaste. Would we be that worried? No, we might think, well, maybe men should see toothpaste ads, not that worried about it. So I do want to emphasize again the idea that it really matters, the outcome really matters. Thank you.

[The prepared statement of Dr. Tucker follows:]

TESTIMONY

COMMITTEE ON ENERGY AND COMMERCE:
SUBCOMMITTEE ON DIGITAL COMMERCE AND CONSUMER
PROTECTION, U.S. HOUSE OF REPRESENTATIVES

CATHERINE TUCKER

November 29, 2017

Executive Summary

This committee hearing is to evaluate the efficacy of current policies and communications with consumers regarding the collection and use of personal data, in the context of the background that algorithms are now often used to determine the content that consumers see and evaluate. My testimony will focus on some of the difficulties of instituting policies surrounding algorithmic bias or fairness, and then talk about some of the unintended trade-offs raised by restrictions of the use and collection of data. To summarize:

- Algorithms may appear biased for many reasons, including economic efficiency. My own research shows that women may be less likely to see an ad for STEM career advice, not because of the usual hypothesized sources of bias, but because other advertisers are willing to pay more for those eyeballs.
- This suggests that, at least in some cases, there may be tradeoffs between correcting bias and economic efficiency when regulating algorithms and their use of data. My prior research suggests that straightforward data usage restrictions impose costs on both firms and consumers.
- In general, identifying an economically optimal approach to data protection is hard because it is difficult to measure what consumers actually want regarding privacy. However, my research suggests that giving consumers a sense of control over how their data is used is welfare-enhancing. Congress should recognize that different types of data have very different types of consequences for consumers, and temper policy to reflect this.

Chairman Latta, Ranking Member Schakowsky, and Members of the Subcommittee: I was honored to receive the invitation to appear before you today to discuss the topic of ‘Algorithms: How Companies’ Decisions About Data and Content Impact Consumers.’”

My name is Catherine Tucker, and I am the Sloan Distinguished Professor of Management at MIT Sloan.

1 Algorithmic Bias or Fairness: The importance of the economic context

Since it is the context of the hearing today, I wanted to start by discussing research I have done into what leads ‘algorithms’ to reach apparently biased results? This was prompted by excellent work done in Computer Science which documented apparent bias in the delivery of internet advertising by algorithms. My recent research has delved into whether there can be reasons grounded in economics that algorithms may appear biased.¹

We ran a field test on Facebook (and replicated on Google and Twitter) which showed that an ad promoting careers in Science, Technology, Engineering and Math (STEM) was shown to between 20-40% more men than women. We then investigate why this occurred:

- It is not because men use these internet sites more than women.
- It is not because women ‘inflict’ this on themselves, by not showing interest or clicking on the ad and the algorithm responds to a perceived lack of interest. If women ever sees the ad, they are more likely than men to click on it.
- It does not seem to echo any cultural bias against women in the workplace. The extent of localized female equality in the workplace is empirically irrelevant for predicting this bias.
- It is instead because other advertise value the opportunity to show ads to female (rather than male) eyeballs. These other advertisers’ willingness to pay more to show ads to

¹See https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2852260 for the full paper.

women, means that an ad that doesn't specify a gender, is shown to fewer women than men. The algorithm is designed to minimize costs, so shows the ad to fewer expensive women than relative cheaper men.

Though this is a case study of a single ad, and a single instance of apparent bias, this research does highlight the following policy insights:

- In this case, it is unlikely that much could have been prevented (or gained) by mandating algorithmic transparency, even supposing it was technologically possible. The apparent bias occurred because of other advertisers' higher valuation of female eyeballs - and this would not have been clear from analyzing an algorithm that was simply intended to minimize costs.
- This bias occurred because of an attempt by the algorithm to minimize costs to advertisers. This opens the possibility that attempts to mandate lack of bias in algorithms can lead to trade-offs if, for example, it prevented all advertisers receiving a 'discount' for showing ads to men. Society may have interest in preventing women from seeing fewer job ads than men, but not in ensuring that women see just as many ads for shoes as men do, and this makes regulating hard.
- It is not clear what the counterfactual would have been. Apparent bias in who sees ads for STEM jobs may happen offline too if employers with job listings shun publications that are more likely to be read by women because ads in such publications are more expensive to advertise in. We only know that this discrepancy occurs online because of the better data and measurement online. This illustrates the importance of knowing the 'but for' world if the algorithm did not exist, but also the difficulties faced in assessing that counterfactual in an offline and less measurable world.

2 Data Protection and Privacy Regulation Tradeoffs

Though it is perhaps stereotypical that an economist would emphasize the need to consider tradeoffs in regulation, I would like to describe some of my recent research which highlights three potential considerations.

2.1 Costs and benefits of privacy regulation

One of the huge benefits of digital data is that it is virtually costless to collect, parse and store. This makes the collection, use and exchange of data for purposes of personalizing the consumer experience both cheaper and easier than a decade ago. However, this lowering of costs has led to evident privacy concerns, as we are now in a world where anyone's data can be viably collated and analyzed by any organization.

One obvious approach in regulation is therefore to simply restrict data collection. As might be expected, such restrictions have real effects on the digital economy which is premised on the use of data. In earlier Congressional testimony I discussed work that I have done into the effects of the EU's e-Privacy Directive which was associated with a 65% decrease in the effectiveness of online advertising for the advertisers I studied.² Similarly, within the US my research has shown that the patchwork of state privacy regulations inhibited the adoption of potentially life-saving digital medical records technology.³

My most recent research has tried to distinguish between the effectiveness of different types of regulation. One recurrent insight has been that rather than simply being focused on imposing costs or restricting flows of data, regulation appears to be more effective when it focuses on restoring a sense of control among consumers. I have found this pattern both

²<https://www.youtube.com/watch?v=meMxH6c1KGE> based on Goldfarb, Avi, and Catherine E. Tucker. "Privacy regulation and online advertising." *Management science* 57.1 (2011): 57-71.

³Miller, Amalia R., and Catherine Tucker. "Privacy protection and technology diffusion: The case of electronic medical records." *Management Science* 55.7 (2009): 1077-1093. and Miller, Amalia R., and Catherine E. Tucker. "Can health care information technology save babies?" *Journal of Political Economy* 119.2 (2011): 289-324.

in responses to very personalized internet advertising,⁴ and also in the realm of personalized medicine and genetic data.⁵ Other researchers have confirmed that the level of perceived control may also positively affect consumer's appreciation of the use of algorithms.⁶

Of course these costs in terms of efficiency need to be set against the potential for benefits for consumers.

2.2 Difficulties in Establishing Consumer Preferences over Data Use

We also ran an experiment which investigated whether undergraduates at MIT would be willing to release what might be considered very personal data regarding their friends' contact information. We found that on average many of them were willing to release the data. There was a subset of students who stated a preference for privacy and did not release the data. However, if this set of students were offered a slice of cheese pizza in exchange for this data, then they were as willing as the rest of the student population to share this information.⁷

There are two ways of interpreting this study. One is that there is often a discrepancy between an individual's privacy preferences as stated in surveys and what they do with their data when faced with very small incentives (or benefits) of giving that data - the so-called privacy paradox. Another is that if MIT students (who I hope are very well informed about data, privacy and algorithms) behave in a way which is so inconsistent with their stated preferences, then we may need more consumer protection. Regardless of interpretation, though, this study emphasizes the extent to which it is hard to use survey-data or stated-preference data to pinpoint exactly what kind of privacy regime might best benefit consumers.

⁴Catherine E. Tucker Social Networks, Personalized Advertising, and Privacy Controls. *Journal of Marketing Research*: October 2014, Vol. 51, No. 5, pp. 546-562.

⁵Miller, Amalia R., and Catherine Tucker. "Privacy Protection, Personalized Medicine, and Genetic Testing." *Management Science* (2017).

⁶Berkeley J. Dietvorst, Joseph Simmons, Cade Massey (2016), Overcoming Algorithm Aversion: People Will Use Algorithms If They Can (Even Slightly) Modify Them, *Management Science*, forthcoming

⁷Athey, Susan, Christian Catalini, and Catherine Tucker. The Digital Privacy Paradox: Small Money, Small Costs, Small Talk. No. w23488. National Bureau of Economic Research, 2017

2.3 Differences in Potential Harm of Data

The other issue I wish to emphasize is that it is easy in a discussion regarding data to treat all the collection and parsing of data as potentially injurious (or not) to consumers.

There are three criteria I use in my own work to consider the potential ‘harm’ of data.⁸

- Could the use of this data lead to negative economic consequences for the consumer?
- For how long could there be potentially negative economic consequences for the consumer associated with this data?
- Could this data also have potential negative consequences for others?

To understand these three criteria, let me contrast two potential types of data: 1) ‘data that I have been searching and researching flowers as a holiday gift for my mother’; and 2) ‘digital genomic data capturing the makeup of my genome.’

The data that I have been browsing for flowers as a holiday gift for my mother is unlikely to have huge economic consequences for me. Instead, the most likely consequences of the release of this data to third parties is that for the next few weeks I receive ads that invite to purchase her flowers, and that may even contain discounts in order to entice me to do so. On the other hand, the public release of my genomic data could lead employers to decide not to employ me if there were reasons to fear for my long term health, and similarly could lead insurance companies to not offer me long term care insurance. Releasing my genomic data has far larger economic consequences.

The data that I have been browsing for flowers as a holiday gift for my mother is unlikely to have much permanent value. I presume there are many people out there who have similarly uninspired gift ideas, so the data is unlikely to have any uniquely identifying value.

⁸See Miller, Amalia R., and Catherine Tucker. “Frontiers of Health Policy: Digital Data and Personalized Medicine.” *Innovation Policy and the Economy* 17.1 (2017): 49-75.

Similarly, the data has little permanent value: In thirty years this data is likely to have little consequence. On the other hand, my genetic data precisely identifies only me, and in thirty years the data will continue to have the same value it has today.

The data that I have been browsing for flowers as a holiday gift, does not really affect anyone else or have informational value about anyone else - except that perhaps it might be possible to piece together my mother's preferred colors. On the other hand my genomic data does have huge spillovers for my siblings, and my children, in that if I am found to be genetically susceptible to something like Huntington's disease, this is a hereditary trait that also elevates their perceived risk levels.

This framework emphasizes that different types of data can have different consequences, and that any regulation, rather than treating all data the same, needs to distinguish between what kinds of data may be actively harmful to consumers and what data may not be.

It also emphasizes that it is tricky to regulate data use by algorithm without consideration as to the economic consequences of the use of that data. There are certain narrow spheres where algorithms and their use of data can have huge consequences, such as employment opportunities and health. However, many uses of algorithms (and data) lead to inconsequential and potentially beneficial increases in personalization of services for consumers and cost-savings for firms.

Thank you for the opportunity to share these thoughts and I look forward to answering your questions.

Mr. LATTA. Thank you very much for your testimony.
And Dr. Pasquale, you are recognized for 5 minutes.

STATEMENT OF FRANK PASQUALE

Mr. PASQUALE. Thank you very much, Chairmen Walden, Blackburn, and Latta and to Ranking Members Schakowsky and Doyle. It is a great honor to be here today.

My testimony is based on my book, "The Black Box Society," in which I distilled about 10 years of research into the role of data and algorithms and argued for the importance of transparency, and I am happy to do that today. I want to argue that the use of data and algorithms by large corporations will be at the core of civil rights, consumer protection, and competition policy for the 21st century. And I will go over each of those and then talk about how this committee can play a role in advancing all three of those goals.

First, with respect to civil rights, I was very glad to hear from Congresswoman Clarke about the letter to Facebook with respect to discriminatory ad profiling. That was discovered last year by ProPublica. There were promises it would be addressed. It was not addressed. And I think that shows some of the failures of self-regulation in the area.

Also in my testimony I talk about racial disparities with respect to ad delivery and disparities with respect to disability status or a health condition. For example, a credit card company deciding to raise the interest rate on someone once they know that the person went for marriage counseling. I think that is a very troubling sort of thing, and we should be able to look into that to get transparency about whether it is happening and to stop it.

Secondly, with respect to consumer protection, Ariel Ezechai and Maurice Stucke are great antitrust law scholars and they say that, given the information asymmetry between large corporations and consumers, consumers now really exist in a Truman Show. It is like a Truman Show online. They know so much about us, we often know so little about their practices, and they show how consumers can be manipulated by data that they don't know about.

So, you know, we may hear a lot about good personalization online, you see things that you want, et cetera, but there is always a dark side to that. There are things, for example, like vulnerability-based marketing, where the marketing could be based on picking out people who are at particularly insecure times in their life or particularly insecure times of day for individuals. And I think this sort of vulnerability-based marketing, predatory loan targeting, all those things are troubling, and not just for traditionally protected groups but also for people, say, in rural areas that might be subject to price discrimination that I discuss in my testimony.

I would also say that with respect to competition, the combination of the power of data in terms of enabling very large digital platforms to decide what consumers see, when they see it, what types of things that they are offered and not offered, that that leads to what I call a self-reinforcing data advantage. What I mean by that is to say that, if you are a large platform, you tend to have more data. When you have more data, you are able to target your

things better to consumers. When you are better able to target to consumers, more consumers come on board.

It is a virtuous cycle in a way, but on the other hand it does risk getting out of hand and creating the types of asymmetries that really you can't overcome as a competitor. And we have seen that, for example, with respect to European action against Google in their antitrust judgment against Google, where they talked about Google potentially privileging its own services over rivals in search results in ways that were opaque to consumers.

And I think that we have got to look at those sorts of dynamics and start to address them. It will be hard, though. And, by the way, I would say that one reason maybe why the U.S. tech scene is doing better than the European one, you know, we have to look at these sort of competitive dynamics, as well, not just regulation. I would also talk about the black box effect here. I would say that it is very hard for us to know exactly what is going on, and we may have only seen the tip of the iceberg here. We may have only scratched the surface.

Now, I have painted a very bleak picture of big data and algorithms in this testimony, but there is good news on the horizon. Over the past decade, a number of visionaries have developed a movement for accountability by users of algorithms. It took a combination of computational, legal, and social scientific skills to unearth each of the examples that I have discussed: troubling collection, bad or biased analysis, or discriminatory use of data. And I hope we talk about all three of those things today.

Empiricists may be frustrated by the black box nature of algorithmic decision making, but they can work with legal scholars and activists if we have freedom of information laws and if we enable people to understand better how data is being collected, how it is being used, how it can lead to discrimination. Journalists also have been teaming up with computer programmers and social scientists to expose new privacy-violating technologies of data collection analysis and use, and they have pushed regulators to crack down on the worst offenders.

I would conclude today by saying that U.S. lawmakers can really help by requiring the openness of algorithms used in many governmental contexts and moving on to empower people to have knowledge of what is going on and how their online lives are being ordered. With that, thank you very much.

[The prepared statement of Mr. Pasquale follows:]

69

Written Testimony of

Frank Pasquale

Professor of Law

University of Maryland

Before the United States House of Representatives

Committee on Energy and Commerce

Subcommittee on Digital Commerce and Consumer Protection

“Algorithms: How Companies’ Decisions About Data and Content Impact Consumers”

November 29, 2017

“Algorithms: How Companies’ Decisions About Data and Content Impact Consumers”Written Testimony of Frank Pasquale¹

I will answer the Committee’s questions in order:

1) How is personal information about consumers collected through the Internet, and how do companies use that information?

Leading digital firms seek out intimate details of customers’ (and potential customers’) lives, but all too often try to give regulators, journalists, and the public at large as little information as they possibly can about their own statistics and procedures.² Internet companies collect more and more data on their users, but tend to fight many of the regulations that would let those same users exercise control over the resulting digital dossiers, and prevent discrimination based on them.

As technology advances, market pressures raise the stakes of the data game. Surveillance cameras become cheaper every year; sensors are embedded in more places.³ Cell phones track our movements; programs log our keystrokes. Intensified data collection promises to make

¹ I wish to thank Sue McCarty and Jennifer Elisa Smith for help in compiling sources on very short notice, and to all those who responded to this request: <https://twitter.com/FrankPasquale/status/935185521080455170>. I was confirmed to testify on November 27 at about ten in the morning, and had to submit this written testimony by 10AM the next day. I therefore ask the reader’s forgiveness for inconsistent footnote formatting and lack of comprehensive coverage of excellent work in algorithmic accountability now being done globally. I have based this testimony, in part, on previous work of mine covering the law and policy of big data, algorithmic accountability, and artificial intelligence.

² Frank Pasquale, *The Black Box Society* (Harvard University Press, 2015).

³ Danielle Citron and Frank Pasquale, *The Scored Society*, *Wash. L. Rev.* (2014)

“quantified selves” of all of us, whether we like it or not.⁴ The resulting information—a vast amount of data that until recently went unrecorded—is fed into databases and assembled into profiles of unprecedented depth and specificity.

But to what ends, and to whose? We are still only beginning to grapple with this problem. Empirical studies may document the value of narrow and particularized forms of profiling. But they only capture small facets of the tip of an iceberg of data use. What lies beneath is hidden via legal measures (such as trade secrecy), physical and administrative safeguards, and obfuscation. A growing algorithmic accountability movement is beginning to expose problems here, but it needs much more support from both government and civil society.⁵

The decline in personal privacy might be worthwhile if it were matched by comparable levels of transparency from corporations and government. But for the most part it is not. Credit raters, search engines, and major banks take in data about us and convert it into scores, rankings, risk calculations, and watch lists with vitally important consequences. But the proprietary algorithms by which they do so are all too often immune from scrutiny.⁶

The personal reputation business is exploding. Having eroded privacy for decades, shady, poorly regulated data miners, brokers and resellers have now taken creepy classification to a

⁴ April Dembosky, “Invasion of the Body Hackers,” *Financial Times*, June 10, 2011; Deborah Lupton, *The Quantified Self* (Polity, 2016); Jenifer S. Winter, “Surveillance in ubiquitous network societies: Normative conflicts related to the consumer in-store supermarket experience in the context of the Internet of Things.” *Ethics and Information Technology*, 16(1), 27-41. doi:10.1007/s10676-013-9332-3.

⁵ Frank Pasquale, *Digital Star Chamber*, at <https://aeon.co/essays/judge-jury-and-executioner-the-unaccountable-algorithm> (2015).

⁶ Cathy O’Neil, *Weapons of Math Destruction* (2016); Frank Pasquale, *Search, Speech, and Secrecy*, at https://ylpr.yale.edu/inter_alia/search-speech-and-secrecy-corporate-strategies-inverting-net-neutrality-debates (2010).

whole new level.⁷ They have created lists of victims of sexual assault, and lists of people with sexually transmitted diseases. Lists of people who have Alzheimer's, dementia and AIDS. Lists of the impotent and the depressed. There are lists of "impulse buyers." Lists of suckers: gullible consumers who have shown that they are susceptible to "vulnerability-based marketing." Even without such inflammatory data, firms can take advantage of unprecedented levels of other data about consumers. The result, as Ryan Calo demonstrates, is that "firms can not only take advantage of a general understanding of cognitive limitations, but can uncover, and even trigger, consumer frailty at an individual level."⁸

The growing danger of breaches challenges any simple attempts to justify data collection in the service of "consumer targeting." Even huge and sophisticated companies can be hacked, and cybercriminals' data trafficking is, unsurprisingly, an obscure topic.⁹ In at least one case, an established U.S. data broker accidentally sold "Social Security and driver's license numbers—as well as bank account and credit card data on millions of Americans" to ID thieves.¹⁰ Until data companies are willing to document and report the precise origins and destinations of all the data they hold, we will never be able to estimate the magnitude of data misuse. Moreover, as the

⁷ Wolfie Christl, *How Companies Use Personal Data Against People: Automated Disadvantage, Personalized Persuasion, and the Societal Ramifications of the Commercial Use of Personal Information* (2017); Theodore Rostow, What Happens When an Acquaintance Buys Your Data?, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2870044 (2016).

⁸ Ryan Calo, Digital Market Manipulation, at http://www.gwlr.org/wp-content/uploads/2014/10/Calo_82_41.pdf; see also Ariel Ezrachi and Maurice Stucke, Is Your Digital Assistant Devious?, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2828117.

⁹ Misha Glenny, *DarkMarket: How Hackers Became the New Mafia* (New York: Vintage Books, 2012) 2 ("this minuscule elite (call them geeks, technos, hackers, coders, securocrats, or what you will) has a profound understanding of a technology that every day directs our lives more intensively and extensively, while most of the rest of us understand absolutely zip about it.").

¹⁰ "Experian Sold Consumer Data to ID Theft Service," *Krebs on Security*, October 20, 2013, <http://krebsonsecurity.com/2013/10/experian-sold-consumer-data-to-id-theft-service/>.

recent Equifax hack showed, massive reservoirs of personal data remain all too vulnerable to misuse.

Even when data is not breached, it can still disadvantage consumers. Think, for example, of the people who type words like “sick,” “stressed,” or “crying” into a search engine or an online support forum and find themselves in the crosshairs of clever marketers looking to capitalize on depression and insecurity.¹¹ Marketers plot to tout beauty products at moments of the day that women feel least attractive.¹² There’s little to stop them from compiling digital dossiers of the vulnerabilities of each of us.¹³ In the hall of mirrors of online marketing, discrimination can easily masquerade as innovation.¹⁴

These methods may seem crude or reductive, but they are beloved by digital marketers. They are fast and cheap and there is little to lose. Once the data is in hand, the permutations are endless, and somebody is going to want them. If you’re a childless man who shops for clothing online, spends a lot on cable TV, and drives a minivan, data brokers may well assume that you

¹¹ Ryan Calo, “Digital Market Manipulation,” at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2309703&download=yes.

¹² PRNewsWire, “New Beauty Study Reveals Days, Times and Occasions When U.S. Women Feel Least Attractive,” October 2, 2013 (news release), <http://www.prnewswire.com/news-releases/new-beauty-study-reveals-days-times-and-occasions-when-us-women-feel-least-attractive-226131921.html>.

¹³ Paul Ohm coined the term “database of ruin” to suggest how damaging information could accumulate about a person. Paul Ohm, “Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization,” *University of California at Los Angeles Law Review* 57 (2010): 1750-51.

¹⁴ Preston Gralla, Opinion, Amazon Prime and the Racist Algorithms, *COMPUTERWORLD* (May 11, 2016, 5:17 AM), <https://www.computerworld.com/article/3068622/internet/amazon-prime-and-the-racist-algorithms.html> (“In Amazon’s mind, race has nothing to do with black neighborhoods being excluded, because no racial demographic data was used in its decision-making. But dig a little deeper, and you’ll see that race has everything to do with it. . . . ‘The Amazon algorithm operates off of an inherited cartography of previous redlining efforts, which created pockets of discrimination, the consequence being that the discrimination continues to be reproduced.’” (quoting Jovan Scott Lewis)).

are heavier than average.¹⁵ And we now know that recruiters for obesity drug trials will happily pay for that analysis, thanks to innovative reporting.¹⁶ But in most cases, we don't know what the owners of massive stores of data are saying about us.

Where does all this data come from? Everywhere. Have you ever searched for “flu symptoms” or “condoms”? That clickstream may be around somewhere, potentially tied to your name (if you were signed in) or the IP address of your computer or perhaps some unique identifier of its hardware.¹⁷ It's a cinch for companies to compile lists of chronic dieters, or people with hay fever. “Based on your credit-card history, and whether you drive an American automobile and several other lifestyle factors, we can get a very, very close bead on whether or not you have the disease state we're looking at,” said a vice president at a company in the health sector.¹⁸ Consumers also worry about the potential misuse of “smart meter” and other technology.¹⁹

¹⁵ Joseph Walker, “Data Mining to Recruit Sick People,” *Wall Street Journal*, December 17, 2013, <http://online.wsj.com/news/articles/SB10001424052702303722104579240140554518458>. The *Journal* tries to explain these Big Data associations by hypothesizing that large men need minivans because they cannot fit into other vehicles. But note how easily we could also rationalize the opposite conclusion: if minivan drivers were pegged as exceptionally fit, we might hypothesize that they used the large vehicle to carry around sports equipment. We should beware post hoc rationalizations of Big Data correlations, particularly when we are unable to review the representativeness of the data processed or the algorithms used to process it.

¹⁶ *Ibid.*

¹⁷ Mary Ebeling, *Health Care and Big Data* (Polity, 2016). Some privacy protective measures are taken with respect to search logs. But, as Nissenbaum and Toubiana observe, “Without an external audit of these search logs, it is currently impossible to evaluate their robustness against de-anonymizing attacks.” V. Toubiana and H. Nissenbaum, “An Analysis of Google Log Retention Policies,” *The Journal of Privacy and Confidentiality* 3, no. 1 (2011): 5. For a search query revelation that proved revealing, despite anonymization efforts, see Thomas Barbaro and Michael Zeller, “A Face Is Exposed for AOL Searcher No. 4417749,” *New York Times*, August 9, 2006, A1.

¹⁸ Walker, “Data Mining to Recruit Sick People.”

¹⁹ Jenifer S. Winter, “(Un)ethical use of smart meters?” In S. Gangadharan (Ed.) *Data and discrimination: Collected essays*. (2014).

Some companies have assembled and sold the mailing addresses and medication lists of depressed people and cancer patients. A firm reportedly combined credit scores and a person's specific ailments into one report.²⁰ The Federal Trade Commission has been trying to nail down a solid picture of these practices,²¹ but exchange of health data is an elusive target when millions of digital files can be encrypted and transmitted at the touch of a button.²² We may eventually find records of data *sales*, but what if it is traded in handshake deals among brokers? A stray flash drive could hold millions of records. It's hard enough for the FTC to monitor America's brick-and-mortar businesses; the proliferation of data firms has completely overtaxed it.²³

Unexpected and troubling uses of data abound. We already know that at least one credit card company has paid attention to certain mental health events, like going to marriage counseling.²⁴ When statistics imply that couples in counseling are more likely to divorce than couples who aren't, counseling becomes a "signal" that marital discord may be about to spill over into financial distress.²⁵ This is effectively a "marriage counseling penalty," and poses a dilemma for policy makers. Left unrevealed, it leaves cardholders in the dark about an important

²⁰ Julie Brill, "Reclaim Your Name," Keynote Address at Computers, Freedom, and Privacy Conference, June 26, 2013. Available at <http://www.ftc.gov/speeches/brill/130626computersfreedom.pdf>.

²¹ FTC, "Data Brokers: A Call for Transparency and Accountability." Federal Trade Commission, May 2014.

²² *Ibid.* ("One health insurance company recently bought data on more than three million people's consumer purchases in order to flag health-related actions, like purchasing plus-sized clothing, the *Wall Street Journal* reported. [The company bought purchasing information for current plan members, not as part of screening people for potential coverage.]")

²³ Peter Maass, "Your FTC Privacy Watchdogs: Low-Tech, Defensive, Toothless," *Wired*, June 28, 2012, <http://www.wired.com/threatlevel/2012/06/ftc-fail/all/>.

²⁴ Charles Duhigg, "What Does Your Credit Card Company Know about You?" *New York Times*, May 17, 2009, <http://www.nytimes.com/2009/05/17/magazine/17credit-t.html?pagewanted=all>. For a compelling account for the crucial role that the FTC plays in regulating unfair consumer practices and establishing a common law of privacy, see Daniel J. Solove and Woodrow Hartzog, "The FTC and the New Common Law of Privacy," *Columbia Law Review* 114 (2014): 583–676.

²⁵ Duhigg, "What Does Your Credit Card Company Know about You?", *New York Times*.

aspect of creditworthiness. Once disclosed, it could discourage a couple from seeking the counseling they need to save their relationship.

There doesn't have to be any established causal relationship between counseling and late payments; correlation is enough to drive action. That can be creepy in the case of objectively verifiable conditions. And it can be devastating for those categorized as "lazy," "unreliable," "struggling," or worse. Runaway data can lead to *cascading disadvantages* as digital alchemy creates new analog realities. Once one piece of software has inferred that a person is a bad credit risk, a shirking worker, or a marginal consumer, that attribute may appear with decision-making clout in other systems all over the economy. There is little in current law to prevent companies from selling their profiles of you.²⁶

Bad inferences are a larger problem than bad data because companies can represent them as "opinion" rather than fact. A lie can be litigated, but an opinion is much harder to prove false; therefore, it is much harder to dispute.²⁷ For example, a firm may identify a data subject not as an "allergy sufferer," but as a person with an "online search propensity" for a certain "ailment or prescription."²⁸ Similar classifications exist for "diabetic-concerned households." It may be easy for me to prove that I don't suffer from diabetes, but how do I prove that I'm not "diabetic-

²⁶ Kashmir Hill, "Could Target Sell Its 'Pregnancy Prediction Score'?" *Forbes*, February 16, 2012, <http://www.forbes.com/sites/kashmirhill/2012/02/16/could-target-sell-its-pregnancy-prediction-score/>.

²⁷ Frank Pasquale, "Reputation Regulation: Disclosure and the Challenge of Clandestinely Commensurating Computing," in *The Offensive Internet: Speech, Privacy, and Reputation*, ed. Saul Levmore and Martha C. Nussbaum (Cambridge, MA: Harvard University Press, 2010), 107–123; Frank Pasquale, "Beyond Innovation and Competition: The Need for Qualified Transparency in Internet Intermediaries," *Northwestern University Law Review* 104 (2010): 105–174.

²⁸ Lois Beckett, "Everything We Know about What Data Brokers Know about You," *ProPublica*, March 7, 2013 (updated September 13, 2013), <http://www.propublica.org/article/everything-we-know-about-what-data-brokers-know-about-you>.

concerned”? And if data buyers are going to lump me in with diabetics anyway, what good does it do me even to bother challenging the record?

Profiling may begin with the original collectors of the information, but it can be elaborated by numerous data brokers, including credit bureaus, analytics firms, catalog co-ops, direct marketers, list brokers, affiliates, and others.²⁹ Brokers combine, swap, and recombine the data they acquire into new profiles, which they can then sell back to the original collectors or to other firms. It’s a complicated picture, and even experts have a tough time keeping on top of exactly how data flows in the new economy.

Most of us have enough trouble keeping tabs on our credit history at the three major credit bureaus. But the Internet has supercharged the world of data exchange and profiling, and Experian, TransUnion, and Equifax are no longer the sole, or even the main, keepers of our online reputations. What will happen when we’ve got dozens, or hundreds, of entities to keep our eyes on?

We’re finding out. They’re already here, maintaining databases that, though mostly unknown to us, record nearly every aspect of our lives. They score us to decide whether we’re targets or “waste,” as media scholar Joseph Turow puts it.³⁰ They keep track of our occupations and preoccupations, our salaries, our home value, even our past purchases of luxury goods.³¹

²⁹ Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers* (Mar. 2012). Available at <http://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf> (providing list of types of data brokers).

³⁰ Joseph Turow, *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth* (New Haven, CT: Yale University Press, 2012).

³¹ Natasha Singer, “Secret E-Scores Chart Consumers’ Buying Power,” *New York Times*, August 18, 2012.

(Who knew that one splurge on a pair of really nice headphones could lead to higher prices on sneakers in a later online search?) There are now hundreds of credit scores for sale, and thousands of “consumer scores,” on subjects ranging from frailty to reliability to likelihood to commit fraud. And there are far more sources of data for all these scores than there are scores themselves.³² Any one of them could change our lives on the basis of a falsehood or a mistake that we don’t even know about.³³

We also need to worry about how public and private databases bleed into one another, potentially reinforcing cycles of disadvantage.³⁴ Such sources can be based on biased data—for example, if police focus their efforts on minority communities, more minorities may end up with criminal records, regardless of whether minorities generally commit more crimes.³⁵ Researchers are revealing that online sources may be just as problematic. As the White House Report on Big Data has found, “big data analytics have the potential to eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education, and the marketplace.”³⁶ Already disadvantaged groups may be particularly hard hit.³⁷

³² Dixon and Gellman, *The Scoring of America*.

³³ Ylan Q. Mui, “Little-Known Firms Tracking Data Used in Credit Scores,” *Washington Post*, July 16, 2011, http://www.washingtonpost.com/business/economy/little-known-firms-tracking-data-used-in-credit-scores/2011/05/24/gIQAXHcWII_print.html. The firm was ChoicePoint (now a part of another, larger firm), a data broker that maintained files on nearly all Americans.

³⁴ Danielle Keats Citron and Frank Pasquale, “Network Accountability for the Domestic Intelligence Apparatus,” *Hastings Law Journal* 62 (2011): 1441–1494; Chris Jay Hoofnagle, “Big Brother’s Little Helpers: How ChoicePoint and Other Commercial Data Brokers Collect, Process, and Package Your Data for Law Enforcement,” *University of North Carolina Journal of International Law and Commercial Regulation* 29 (2004): 595–638; Jon D. Michaels, “All the President’s Spies: Private–Public Intelligence Partnerships in the War on Terror” (2008).

³⁵ Associated Press, “EEOC Sues over Criminal Background Checks,” *CBSNews*, June 11, 2013, http://www.cbsnews.com/8301-505123_162-57588814/eoc-sues-over-criminal-background-checks/.

³⁶ Executive Office of the President, *Big Data: Seizing Opportunities, Preserving Values* (2014).

For example, consider one computer scientist's scrutiny of digital name searches. In 2012, Latanya Sweeney, former director of the Data Privacy Lab at Harvard and now a senior technologist at the Federal Trade Commission, suspected that African Americans were being unfairly targeted by an online service. When Sweeney searched her own name on Google, she saw an ad saying, "Latanya Sweeney: Arrested?" In contrast, a search for "Tanya Smith" produced an ad saying, "Located: Tanya Smith."³⁸ The discrepancy provoked Sweeney to conduct a study of how names affected the ads served. She suspected that "ads suggesting arrest tend to appear with names associated with blacks, and neutral ads or no [such] ads tend to appear with names associated with whites, regardless of whether the company [purchasing the ad] has an arrest record associated with the name." She concluded that "Google searches for typically African-American names lead to negative ads posted by [the background check site] InstantCheckmate.com, while typically Caucasian names draw neutral ads."³⁹

After Sweeney released her findings, several explanations for her results were proposed. Perhaps someone had deliberately programmed "arrest" results to appear with names associated with blacks? That would be intentional discrimination, and Instant Checkmate and Google both vehemently denied it. On the other hand, let us suppose that (for whatever reasons) web

³⁷ David Talbot, "Data Discrimination Means the Poor May Experience a Different Internet," *Technology Review*, Oct. 9, 2013, at <http://www.technologyreview.com/news/520131/data-discrimination-means-the-poor-may-experience-a-different-internet/> (discussing work of Kate Crawford and Jason Schultz).

³⁸ Devony B. Schmidt, "Researchers Present Findings on Online Criminal Record Websites," *The Harvard Crimson*, November 20, 2012, <http://www.thecrimson.com/article/2012/11/20/research-finds-profiling/>.

³⁹ Latanya Sweeney, "Discrimination in Online Ad Delivery," *Communications of the ACM* 56 (2013): 44. She ultimately found "statistically significant discrimination in ad delivery based on searches of 2184 racially associated personal names," in that ads suggesting arrest (as in the question, Arrested?) were likely to appear in the context of names associated with blacks even when there was no actual arrest record associated with the name. This was not true of names associated with whites. There are many more examples of very troubling, racially charged sorting in Safiya U. Noble, *Algorithms of Oppression* (forthcoming, 2018).

searchers tended to click on Instant Checkmate ads more often when names associated with blacks had “arrest” associations, rather than more neutral ones. In that case, the programmer behind the ad-matching engine could say that all it is doing is optimizing for clicks—it is agnostic about people’s reasons for clicking.⁴⁰ It presents itself as a cultural voting machine, merely registering, rather than creating, perceptions.⁴¹

Given algorithmic secrecy, it’s very hard to know exactly what’s going on here.⁴²

Perhaps a company had racially inflected ad targeting; perhaps Sweeney’s results arose from other associations in the data.⁴³ But without access to the underlying coding and data, it is very difficult to adjudicate the dispute. That is troubling, because as FTC chair Edith Ramirez has argued, we must “ensure that by using big data algorithms they are not accidentally classifying

⁴⁰ “Racism Is Poisoning Online Ad Delivery, Professor Says,” *MIT Technology Review*, February 4, 2013, <http://www.technologyreview.com/view/510646/racism-is-poisoning-online-ad-delivery-says-harvard-professor/>.

⁴¹ Toon Calders & Indre Zliobaite, “Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures,” in *Discrimination and Privacy in the Information Society* (Bart Custers, et al., eds.) (Heidelberg: Springer, 2013).

⁴² Trade secrecy will likely continue to blunt efforts to get to the bottom of issues like the ones identified by Sweeney. However, there are forms of auditing that can help us understand what is going on in automated systems without full transparency of data or algorithms. See, e.g. Christian Sandvig et al., *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms*, at <http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf>; Sandra Wachter, Brent Mittelstadt, and Chris Russell, *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3063289.

⁴³ On the question of attribution and intent, see Frank Pasquale, *Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society*, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3002546; Luciano Floridi, *Faultless responsibility: on the nature and allocation of moral responsibility for distributed moral actions*, at <https://www.ncbi.nlm.nih.gov/pubmed/28336791>.

people based on categories that society has decided—by law or ethics—not to use, such as race, ethnic background, gender, and sexual orientation.”⁴⁴

2) How do companies make decisions about content that consumers see online?

The same problems of opacity that plague the dark market in personal data, also afflict online content display and ordering. A large platform may marginalize (or entirely block) potential connections between audiences and speakers. Consumer protection concerns arise, for platforms may be marketing themselves as open, comprehensive, and unbiased, when they are in fact closed, partial, and self-serving. Responding to protests, accused platforms have tended both to assert a right to craft the information environments they desire, and to abjure responsibility, claiming to merely reflect the desires and preferences of the user base. Such contradictory responses betray an opportunistic commercialism at odds with the platforms’ touted social missions. Large platforms should be developing (and holding themselves to) more ambitious standards, rather than warring against privacy, competition, and consumer protection laws.⁴⁵ These regulations enable a more vibrant public sphere. They also defuse the twin specters of monopolization and total surveillance, which are grave threats to freedom of expression.

Policymakers should also consider expanding some core principles of network neutrality beyond the physical layer of the internet to very large enterprises at the social, search, and app level.⁴⁶ Bottlenecks can threaten competition at any layer of the network.

⁴⁴ Ibid.

⁴⁵ Frank Pasquale, Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2779270 (2016).

⁴⁶ Frank Pasquale, Internet Nondiscrimination Principles: Commercial Ethics for Carriers and Search Engines, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1134159 (2008).

Renewed enforcement of anti-discrimination law is also critical in online contexts. One thing is clear: self-regulation is not working. As reported in ProPublica, after Facebook was caught enabling discriminatory housing ads online in 2016, it pledged to change its system to fix the problem. But the issue persists.⁴⁷

The interaction between paid and organic search results also merits attention here.⁴⁸ Google's misadventures in the medical space suggest some of the problems that can arise when automated systems are not up to the tasks that they have taken on. According to a recent report, its neglect enabled predatory addiction clinics to displace more established ones, and may be making discrimination as to source of insurance coverage all too easy.⁴⁹ As a de facto addiction center referral center, it has effectively let bad actors game its systems. The company may plead that it is not responsible. But one has to wonder about whether its extraordinarily high profit levels are premised on a level of neglect of the vulnerable that is unacceptable.⁵⁰ An insurer that

⁴⁷ Julia Angwin, Ariana Tobin and Madeleine Varner, Facebook (Still) Letting Housing Advertisers Exclude Users by Race, at <https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin>. Angwin's series of articles on algorithmic bias at ProPublica, as well as her earlier "What They Know" series in the Wall Street Journal, are a vital resource for those interested in online discrimination. *What They Know*, at <http://www.wsj.com/public/page/what-they-know-digital-privacy.html>.

⁴⁸ For an account of extant regulation, see Frank Pasquale, *Beyond Innovation and Competition*, Northwestern L. Rev. (2010) (discussing the FTC's sponsorship disclosure guidelines); Danny Sullivan, A Letter To The FTC Regarding Search Engine Disclosure Compliance, at <https://searchengineland.com/a-letter-to-the-ftc-regarding-search-engine-disclosure-124169> (discussing the need to ensure that FTC guidelines on sponsorship disclosure are actually enforced).

⁴⁹ Cat Ferguson, How Disreputable Rehabs Game Google to Profit off Patients, The Verge, at <https://www.theverge.com/2017/9/7/16257412/rehabs-near-me-google-search-scam-florida-treatment-centers>; David Dayen, Google is So Big, It is Now Shaping Policy to Combat the Opioid Epidemic—And Screwing it Up, The Intercept, at <https://theintercept.com/2017/10/17/google-search-drug-use-opioid-epidemic/>.

⁵⁰ Will Oremus, *Facebook's Broken Promises*, SLATE (Nov. 24, 2017, 9:47 AM), http://www.slate.com/articles/technology/technology/2017/11/why_facebook_broke_its_promise_to_stop_allowing_racist_housing_ads.html ("fixing these problems requires time, resources, and, yes, manpower—all of which not only cut into Facebook's profits but run counter to its entire culture and

maintained networks of manifestly incompetent or unqualified professionals could be either secondarily or directly liable for its failures. An online intermediary irresponsibility lobby has worked hard to entrench ever more expansive readings of Section 230 of the Communications Decency Act in order to immunize firms like Google from such responsibility. At some point, though, the collateral consequences of such policies need to be taken into account.⁵¹

The same concerns also arise in education and finance. As Sam Adler-Bell explains, “Debt relief companies are counting on you doing what most people do when a serious and complicated problem strikes: Google it. . . . [T]he CFPB [has] sent letters to Microsoft, Google, Facebook, and Yahoo warning them that student debt scammers were using their ad services and search products to ‘lure distressed borrowers.’”⁵² The college classroom itself may be stratified by big data in ways that are hard for students to fully understand.⁵³ Librarians and information science professionals are exposing the stakes of different algorithmic systems of ordering

philosophy.”). On intermediary irresponsibility generally, see Frank Pasquale, *The Automated Public Sphere*, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3067552 (2017).

⁵¹ There are some exceptions to talismanic 230 immunities. See, e.g., Jake Pearson, *How a Career Con Man Led a Federal Sting that cost Google \$500 million*, at <https://www.wired.com/2013/05/google-pharma-whitaker-sting/> (“As part of the agreement, the company acknowledged that it had helped presumably Canadian online pharmacies use AdWords as early as 2003, that it knew US customers were buying drugs through these ads, that advertisers were selling drugs without requiring prescriptions, and that Google employees actively helped advertisers circumvent their own pharmaceutical policies and third-party verification services.”).

⁵² Sam Adler-Bell, *Scam Artists are preying on Student Debt Holders – and Google is Helping*, COMMENTARY: THE CENTURY FOUNDATION (Sept. 14, 2015), <https://tcf.org/content/commentary/scam-artists-are-preying-on-student-debt-holders-and-google-is-helping/>.

⁵³ Frank Pasquale, *Big Data: It’s Worse than you Thought*, at <http://www.latimes.com/opinion/op-ed/la-oe-0116-pasquale-reputation-repair-digital-history-20150116-story.html> (“colleges are now using data to warn professors about at-risk students. Some students arrive in the classroom with a “red light” designation — which they don’t know about, and which is based on calculations they can’t access”).

information.⁵⁴ And even at the earliest stages of education, algorithmic mediation is having widespread (and largely unexamined) effects.⁵⁵

Ariel Ezrachi and Maurice Stucke have described the resulting online landscape as a version of the movie *The Truman Show*, where we are constantly manipulated in ways we can neither fully anticipate nor guard against.⁵⁶ Many users have little appreciation of the way that algorithms are shaping their online experience.⁵⁷ For example, Navneet Alang has reported that Amazon “uses AI to push customers to higher-priced products that come from preferred partners.”⁵⁸ These methods may be becoming more widespread.⁵⁹ In their account of the “algorithmic consumer,” Michael S. Gal & Niva Elkin-Koren conclude that:

⁵⁴ See, e.g., Algorithmic Bias in Library Discovery Systems, at <https://matthew.reidsrow.com/articles/173>; Moritz Hardt, How big data is unfair: Understanding unintended sources of unfairness in data driven decision making, at <https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de>.

⁵⁵ Elana Zeide, The Structural Consequences of Big Data-Driven Education (June 23, 2017). Big Data, Vol 5, No. 2 (2017): 164-172, at <https://ssrn.com/abstract=2991794> (“[B]ig data-driven tools define what ‘counts’ as education by mapping the concepts, creating the content, determining the metrics, and setting desired learning outcomes of instruction. These shifts cede important decision-making to private entities without public scrutiny or pedagogical examination. In contrast to the public and heated debates that accompany textbook choices, schools often adopt education technologies ad hoc.”).

⁵⁶ Ben Schiller, You Are Being Exploited By The Opaque, Algorithm-Driven Economy, at <https://www.fastcompany.com/40447841/you-are-being-exploited-by-the-opaque-algorithm-driven-economy>.

⁵⁷ Motahhare Eslami et al., “I Always Assumed That I Wasn’t Really That Close to [Her]”: Reasoning About Invisible Algorithms in the News Feed, 2015 PROC. 33RD ANN. ACM CONF. ON HUM. FACTORS COMPUTING SYS. 153, available at: http://www-personal.umich.edu/~csandvig/research/Eslami_Algorithms_CHI15.pdf (Study focused on user engagement with Facebook’s News Feed algorithm, finding “that 62.5% of participants were not aware of the algorithm’s existence.”).

⁵⁸ Navneet Alang, *Turns Out Algorithms are Racist*, NEW REPUBLIC (Aug. 31, 2017), <https://newrepublic.com/article/144644/turns-algorithms-racist>, citing Julia Angwin & Surya Mattu, Amazon Says It Puts Customers First. But Its Pricing Algorithm Doesn’t, PROPUBLICA (Sept. 20, 2016, 8:00 AM), <https://www.propublica.org/article/amazon-says-it-puts-customers-first-but-its-pricing-algorithm-doesnt>).

⁵⁹ Katie Pedersen, Greg Sadler and Virginia Smart, *How Companies Use Personal Data to Charge Different People Different Prices for the Same Product*, CBC NEWS (Nov. 24, 2017, 2:21 PM),

[V]ulnerability to biases and errors embedded in the code or drawn from the data is not easily overcome. A consumer who is unaware of such assumptions will likely also be unaware of any choices she has forgone. This type of failure, involving unknown unknowns, is likely to be difficult to fix. Consumers may find it increasingly difficult — or not worth their time — to exercise oversight over sophisticated and opaque systems.⁶⁰

Rural or socioeconomically disadvantaged areas may be hardest hit.⁶¹ Moreover, many consumers may not even believe they have to guard against price discrimination, because they assume it is illegal.⁶² Or they may find it futile to even try to protect themselves against that and other forms of discrimination, given the opacity of contemporary data practices.⁶³ Fortunately,

<http://www.cbc.ca/news/business/marketplace-online-prices-profiles-1.4414240>; *Price-bots Can Collude Against Consumers*, THE ECONOMIST: FREE EXCHANGE BLOG (May 6, 2017), <https://www.economist.com/news/finance-and-economics/21721648-trustbusters-might-have-fight-algorithms-algorithms-price-bots-can-collude>.

⁶⁰ Michael S. Gal & Niva Elkin-Koren, *Algorithmic Consumers*, 30 HARV. J.L. & TECH. 309 (2017), available at: <http://jolt.law.harvard.edu/assets/articlePDFs/v30/30HarvJLTech309.pdf>.

⁶¹ Jennifer Valentino-DeVries, Jeremy Singer-Vine & Ashkan Soltani, *Websites Vary Prices, Deals Based on Users' Information*, WALL ST. J. (Dec. 24, 2012), <https://www.wsj.com/articles/SB1000142412788732377204578189391813881534> (“using geography as a pricing tool can also reinforce patterns that e-commerce had promised to erase: prices that are higher in areas with less competition, including rural or poor areas. It diminishes the Internet's role as an equalizer.”); Kaveh Waddell, *The Internet May Be as Segregated as a City*, THE ATLANTIC (Sept. 6, 2016), <https://www.theatlantic.com/technology/archive/2016/09/the-internet-may-be-as-segregated-as-a-city/498608/>.

⁶² Neil Howe, *A Special Price Just for You*, FORBES (Nov. 17, 2017, 5:56 PM), <https://www.forbes.com/sites/neilhowe/2017/11/17/a-special-price-just-for-you/#dfd7bce90b32> (“When consumers realize that price discrimination is occurring, they object. Most, in fact, mistakenly believe it to be illegal. A 2005 Annenberg Center study found that 64% of adult Internet users thought it was illegal for e-commerce sites to charge different prices to different customers—and 71% thought it was illegal for brick-and-mortar retailers to do so.”).

⁶³ Mary Madden, Michele Gilman, Karen Levy & Alice Marwick, *Privacy, Poverty, and Big Data: A Matrix of Vulnerabilities for Poor Americans*, 95 WASH. U.L.R. 53 (2017), at https://openscholarship.wustl.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=6265&context=law_lawreview (“In cases of big-data-related decision-making and discrimination,

researchers are now documenting algorithmic biases to raise public awareness of them.⁶⁴ But this is a problem that individuals, on their own, cannot hope to solve. It has to be addressed by policymakers.

There is a range of responses that policymakers should look into.⁶⁵ Since at least 2008, scholars have proposed new agencies to ensure algorithmic fairness and accountability.⁶⁶ Some researchers argue for a “watchdog system that allows users to detect discriminatory practices.”⁶⁷ Joanna Bryson has proposed that “Citizens (or perhaps citizens’ advocates, see next paragraph) should be able to trigger audits of software systems when they suspect conditions such as a) the inappropriate or unauthorized use of data, or b) unfair or unlawful bias.”⁶⁸ Whatever the details, one thing is clear: algorithmic “pricing may require new approaches to competition investigations, and possibly even to the legal definition of competition infringements,” as well as

it is nearly impossible for respondents to know what personal or behavioral information may have factored into an unfavorable outcome.”).

⁶⁴ Jerry Useem, How Online Shopping Makes Suckers of Us All, *The Atlantic* (May 2017), <https://www.theatlantic.com/magazine/archive/2017/05/how-online-shopping-makes-suckers-of-us-all/521448/>.

⁶⁵ Jędrzej Niklas, The regulatory future of algorithms, at <http://blogs.lse.ac.uk/mediapolicyproject/2017/08/15/the-regulatory-future-of-algorithms/>.

⁶⁶ Oren Bracha & Frank Pasquale, Federal Search Commission? Access, Fairness, And Accountability in the Law of Search, at <http://www.lawschool.cornell.edu/research/cornell-law-review/upload/Bracha-Pasquale-Final.pdf>; Andrew Tutt, An FDA for Algorithms, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2747994; Ryan Calo, The case for a federal robotics commission, at <https://www.brookings.edu/research/the-case-for-a-federal-robotics-commission/>.

⁶⁷ Jakub Mikians, Laszlo Gyarmati, Vijay Erramelli & Nikolaos Laoutaris, Detecting Price and Search Discrimination on the Internet, 2012 Proc. 11th ACM Workshop On Hot Topics Networks 79, at http://www.ccs.neu.edu/home/cbw/static/class/5750/papers/hotnets2012_pd_cr.pdf.

⁶⁸ Bryson, Testimony for the The House of Lords Select Committee on Artificial Intelligence, at <https://joanna-bryson.blogspot.com/2017/09/testimony-for-the-house-of-lords-select.html>.

new consumer protections.⁶⁹ As Rick Swedloff has argued, “while big data may be a natural next step in risk classification, it may require a revolutionary approach to regulation.”⁷⁰

3) How effective are current policies and communications with consumers regarding the collection and use of personal data?

Current policies are failing because, when it comes to consumers’ relationships with dominant providers, they are based on a category mistake. Online “terms of service” are not ordinary contracts. They cannot be negotiated or otherwise altered. They are take-it-or-leave-it deals offered by must-have services.⁷¹ Thus privacy policies are experienced, by most, as a form of “privacy theater,” and may even be viewed as “exposure policies,” since they so often reserve so many rights to data exploitation to the more powerful entity in the so-called bargain.

This category mistake arose out of a naïvely economic approach to privacy as a normal good or service to be bargained for, like any other. Within a neoclassical economic framework, the relationship between Internet privacy and competition is direct and positive. Consumers set out to obtain an optimal amount of privacy as a feature of the Internet services they consume. Just as a car buyer might choose a Volvo over a Ford because the Volvo is said to have better crash impact protection than the Ford, so too might a search engine user choose DuckDuckGo over Google because of the privacy DuckDuckGo offers.⁷² Companies compete to offer more or

⁶⁹ Oxera Economic Council, *When Algorithms Set Prices: Winners And Losers* (2017), at https://www.regulation.org.uk/library/2017-Oxera-When_algorithms_set_prices-winners_and_losers.pdf.

⁷⁰ Swedloff, *Risk Classification's Big Data (R)evolution*, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2566594 (2014).

⁷¹ The rest of this section is drawn from Frank Pasquale, *Privacy, Antitrust, and Power*, *George Mason L. Rev.* (2013).

⁷² Google’s advocates frequently mention DuckDuckGo as a competitor, but industry experts are skeptical. Brooke Gladstone, *Can a Small Search Engine Take on Google?*, *ON THE MEDIA*, at <http://www.onthemedial.org/2013/apr/12/duck-duck-go-and-competition-search-market/transcript/>, Apr.

less privacy to users. If there are many companies in a given field, they will probably offer many different levels of privacy to consumers. If consumers choose to use services from companies that offer little to no privacy protection, that reveals a preference to spend little to nothing on (or looking for) privacy.

Within the neoclassical model, there is little reason for government to limit a firm's collection, analysis, and use of data. Consumers individually decide how much information they want to release about themselves into commercial ecosystems. Indeed, such limits might even undermine the competition that is supposed to be the primary provider of privacy.⁷³ Companies may need to share data with one another in order to compete effectively. Privacy laws that interfere with that sharing press firms to merge, so that they can seamlessly utilize data that they would have sold or traded to one another in the absence of privacy laws restricting that action.

It would be nice to believe that market forces are in fact promoting optimal levels of privacy. It would also be comforting if antitrust law indirectly promoted optimal privacy options by assuring a diverse range of firms that can compete to supply privacy at various levels (and in

12, 2013 ("DuckDuckGo doesn't collect any of your personal data, at all, full stop. . . . Still, Danny Sullivan, who founded Search Engine Land.com, laughed when Google cited DuckDuckGo as a contender. 'It would be like a major baseball player saying, yeah, there's plenty of great athletes out there, look at this kid who's in eighth grade. And the only reason it can really get counted is because there's relatively little competition in the space'" [said Sullivan]). Sullivan's points here were prophetic, and likely only to become more so.

⁷³ Randal C. Picker, *Competition and Privacy in Web 2.0 and the Cloud*, 103 NW. U. L. REV. COLLOQUY 1, 11–12 (2008) ("An uneven playing field that allows one firm to use the information that it sees while blocking others from doing the same thing creates market power through limiting competition. We rarely want to do that. And privacy rules that limit how information can be used and shared across firms will artificially push towards greater consolidation, something that usually works against maintaining robust competition."). Picker argued that privacy laws restricting interfirm (but not intrafirm) data-sharing may actually undermine competition by encouraging consolidation of firms..

various forms).⁷⁴ But this position is not remotely plausible. Antitrust law has been slow to recognize privacy as a dimension of product quality, and the competition that antitrust promotes can do as much to trample privacy as to protect it.⁷⁵ In an era of big data, every business has an incentive to be nosy in order to maximize profits.⁷⁶

This account of “competition promoting privacy” only achieves surface plausibility by privileging the short-term “preferences” of consumers to avoid data sharing.⁷⁷ The narrowness of “notice-and-consent” as a privacy model nicely matches the short-term economic models now dominating American antitrust law. The establishment in the field is largely unconcerned with too-big-to-fail banks, near monopoly in search advertising, media consolidation, and other forms of industrial concentration. By focusing myopically on efficiency gains that can be temporary or exaggerated, they gloss over the long term pathologies of corporate concentration.⁷⁸ So, too, does a notice-and-consent privacy regime privilege on-the-fly, snap judgments of consumers to

⁷⁴ “Indirectly” is used here because it is now antitrust orthodoxy that this field of law exists only to protect competition, not competitors, and therefore is concerned first and foremost with *directly promoting consumer welfare*. For an account of the rise of consumer welfare as antitrust’s standard (and the problems this has caused), see Barak Orbach, *How Antitrust Lost Its Goal*, 81 *FORDHAM L. REV.* 2253, 2253 (2013) (“while ‘consumer welfare’ was offered as a remedy for reconciling contradictions and inconsistencies in antitrust, the adoption of the consumer welfare standard sparked an enduring controversy, causing confusion and doctrinal uncertainty.”).

⁷⁵ As Paul Ohm has documented, competition among broadband ISPs has led them to “search for new sources of revenue . . . [by] ‘trading user secrets for cash,’ which Google has proved can be a very lucrative market.” Paul Ohm, *The Rise and Fall of Invasive ISP Surveillance*, 2009 *U. ILL. L. REV.* 1417, 1420 (2009) (describing the many commercial pressures leading carriers to “monetize[] behavioral data at the expense of user privacy”).

⁷⁶ VIKTOR-MAYER SCHONBERGER AND VICTOR CUKIER, *BIG DATA* (2013).

⁷⁷ Even if consumers tried to opt out more often, notice-and-consent is increasingly irrelevant because, in an era of big data, whatever one might try to hide by keeping certain pieces of data private is increasingly easy to infer from other pieces of data. *Id.*

⁷⁸ For a critique of contemporary antitrust, see BARRY C. LYNN, *CORNERED: THE NEW MONOPOLY CAPITALISM AND THE ECONOMICS OF DESTRUCTION* 30 (2010) (“superconsolidation is pretty much standard operating procedure for all industries in the United States these days.”); Frank Pasquale, *When Antitrust Becomes Protrust*, at <https://www.competitionpolicyinternational.com/wp-content/uploads/2017/05/CPI-Pasquale.pdf>.

“opt-in” to one-sided contracts, over a reflective consideration of how data flows might be optimized for consumers’ interests generally. As privacy declines and companies consolidate, mainstream antitrust and privacy theory often legitimates the process. Some scholarship can even amount to the “structural production of ignorance,” characterizing scenarios as “consent” and “competition” when they are experienced by consumers and users as coercive and monopolistic.⁷⁹

In response to these problems, many advocates have called for more transparency. Privacy regulators should also require auditors to gain a deep understanding of data broker practices, so they can quickly detect and deter failures to adhere to data collection, labeling, and filtering standards. The key here is to begin separating out the many zones of life Big Data grandees are so keen to integrate in databases. Health privacy experts have already spearheaded “data segmentation for privacy” in medical records, allowing for, say, a person to segregate entries from a psychiatrist from those coming from a podiatrist. It is time for the controllers of Big Data generally to become far more careful about how they log data, to be sure its collection, analysis, and use can be influenced by public values, and not just the profit motive.⁸⁰

⁷⁹ Robert N. Proctor, *Agnotology: A Missing Term to Describe the Cultural Production of Ignorance (and Its Study)*, in *AGNOTOLOGY: THE MAKING AND UNMAKING OF IGNORANCE 3* (Robert N. Proctor & Londa N. Schiebinger eds., 2008).

⁸⁰ Recent rules proposed in New York in the wake of the Equifax scandal may also be of use here. See https://www.governor.ny.gov/sites/governor.ny.gov/files/atoms/files/DFS_CRA_Reg.pdf#_blank; see generally *Written Testimony of Frank Pasquale Before the United States Senate Committee on Banking, Housing, and Urban Affairs*, “Exploring the Fintech Landscape,” at https://www.banking.senate.gov/public/_cache/files/0a92ad09-6834-4d7e-901a-6ae5c51572ae/6F5BB3DB26E6C8891F7A5627A3678DCE.pasquale-testimony-9-12-17.pdf; *Testimony and Statement for the Record of Marc Rotenberg, Hearing on Consumer Data Security and the Credit Bureaus Before the Committee on Banking, Housing, and Urban Affairs of the United States Senate*, at https://www.banking.senate.gov/public/_cache/files/19fa71b4-224a-4331-aec7-2fc99081e383/FF627C28C101D75E809511A6D36B284B.rotenberg-testimony-10-17-17.pdf.

4) Conclusion

I have painted a bleak picture of big data and algorithms in this testimony. However, there is good news on the horizon. Over the past decade, a number of visionaries have developed a movement for accountability by the users of algorithms.⁸¹ It took a combination of computational, legal, and social scientific skills to unearth each of the examples discussed above – troubling collection, bad or biased analysis, and discriminatory use.⁸² Empiricists may be frustrated by the ‘black box’ nature of algorithmic decision-making; they can work with legal scholars and activists to open up certain aspects of it (via freedom of information and fair data practices). Journalists, too, have been teaming up with computer programmers and social scientists to expose new privacy-violating technologies of data collection, analysis, and use – and to push regulators to crack down on the worst offenders.

Researchers are going beyond the analysis of extant data, and joining coalitions of watchdogs, archivists, open data activists, and public interest attorneys, to assure a more balanced set of ‘raw materials’ for analysis, synthesis, and critique. Social scientists and others must commit to the vital, long term project of assuring that algorithms are producing fair and relevant documentation; otherwise large internet firms, states, banks, insurance companies and other powerful actors will make and own more and more inaccessible data about society and people. Algorithmic accountability is a big tent project, requiring the skills of theorists and practitioners, lawyers, social scientists, journalists and others. It’s an urgent, global cause with

⁸¹ Groups like AINow, Data & Society, Data for Black Lives, and many others are part of this trend. Early scholarly work included Lucas Introna & Helen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters* (2000), at <https://www.nyu.edu/projects/nissenbaum/papers/ShapingTheWeb.pdf>.

⁸² This section is largely drawn from Frank Pasquale, *Digital Star Chamber*, at <https://aeon.co/essays/judge-jury-and-executioner-the-unaccountable-algorithm>.

committed and mobilized experts looking for support. Lawmakers can help by, for example, requiring openness in algorithm used in many governmental contexts.⁸³

The world is full of algorithmically driven decisions. One errant or discriminatory piece of information can wreck someone’s employment or credit prospects. It is vital that citizens be empowered to see and regulate the digital dossiers of business giants and government agencies.⁸⁴ Even if one believes that no information should be ‘deleted’ – that every slip and mistake anyone makes should be on a permanent record for ever – that still leaves important decisions to be made about the processing of the data. Algorithms can be made more accountable, respecting rights of fairness and dignity for which generations have fought. The challenge is not technical, but political, and the first step is law that empowers people to see and challenge what algorithms are saying about us.

⁸³ See, e.g., the proposal *A Local Law to amend the administrative code of the city of New York, in relation to automated processing of data for the purposes of targeting services, penalties, or policing to persons*, at <http://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0&Options=&Search=>; Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker, and Kate Crawford, *AI Now 2017 Report* (“Core public agencies, such as those responsible for criminal justice, healthcare, welfare, and education (e.g. “high stakes” domains) should no longer use “black box” AI and algorithmic systems”). I expect many of these algorithms to undergo increasing scrutiny in coming years. See, e.g., Virginia Eubanks, *Automated Inequality* (forthcoming, 2018).

⁸⁴ European data protection law should provide some inspiration for US policymakers as well here. See, e.g., Andrew D. Selbst and Julia Powles, *Meaningful Information and the Right to Explanation*, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3039125; Gianclaudio Malgieri Giovanni Comandé, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, at <https://academic.oup.com/idpl/advance-article-abstract/doi/10.1093/idpl/ix019/4626991>. For background on the development of “explainable AI,” see Cliff Kuang, *Can an AI Be Taught to Explain Itself?*, at <https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html>. Policymakers should try to channel the development of AI that ranks, rates, or sorts humans, toward explainable (rather than black box) models.

Mr. LATTA. Thank you for your testimony this morning.
And Dr. Kearns, you are recognized for 5 minutes.

STATEMENT OF MICHAEL KEARNS

Dr. KEARNS. Thank you. Chairmen Blackburn and Latta, Ranking Members Doyle and Schakowsky, and other distinguished members of the subcommittees, thank you for the opportunity to testify at this important hearing. My name is Michael Kearns, and I am a computer and information science professor at the University of Pennsylvania. I am an active researcher in the field of machine learning, and I have consulted extensively on the use of machine learning in the technology and finance industries.

The fields of machine learning and artificial intelligence now play a central role in virtually every sector in which large data sets are present. The number of instances in which the use of machine learning has provided tangible societal benefits, such as in medical diagnosis, is large and growing. Machine learning also increasingly plays a central role in the data collection and use practices of consumer-facing technology companies.

Today I want to discuss data intimacy, which is the notion that machine learning enables companies to routinely draw predictions and inferences about users that go far deeper than the apparent face value of the data collected as part of online activities. It is not simply a question of whether consumer-facing tech companies are collecting large volumes of data, such companies are collecting information that provides or allows inferences regarding intimate details about our personal lives.

Search engine queries permit inferences about our physical, financial, and psychological conditions. Social media users routinely reveal intimate opinions, beliefs, or affiliations. For example, a recent study showed that using machine learning, anonymous social relationship data permitted accurate identification of romantic partners for over 55 percent of users. Another study concluded that Facebook's algorithms and models are capable of identifying social relationships of which its users are themselves unaware. And religious and political beliefs can be accurately predicted from apparently unrelated social search and shopping activity.

Consumer-facing tech companies in the United States have amassed an almost unimaginable set of data about consumers, which enables machine learning and artificial intelligence to make predictions and inferences about consumer behavior and preferences. These large and diverse data sets are the foundation for effective algorithms and models, and companies compete vigorously to amass or acquire these data sets. For example, search engines provide vast amounts of data about consumers' interests in the manner in which they conduct searches. Similarly, mobile operating system data provides a treasure trove of information regarding virtually everything a consumer does on a mobile device as well as their physical location.

In addition to knowing with whom a consumer affiliates directly, social media platforms are able to accumulate information about who a consumer follows or what he or she likes. However, while the quantity of data is critical to develop accurate algorithms and models, the quality and intimacy of such data is equally or more

important in discerning consumer preferences and behaviors. Increasingly, machine-learning-based algorithms are utilized not only to determine consumer purchasing habits, but also to infer a consumer's emotions, moods, and mental states.

While machine learning is employed most commonly and pervasively to target advertising as we have seen in the media recently, algorithms can also be utilized to generate or incite certain emotional responses. From a privacy perspective, perhaps the most important overarching conclusion is that the intimacy of consumer data cannot be measured by metrics that fail to account for the nature, diversity, and content of the data and, most importantly, its potential uses for modeling and inferences.

It is both common and possible that the highest-volume data sources can reveal little about the consumers who generate that traffic, whereas more specialized data can directly and indirectly reveal the most private and personal details about consumers. In fact, the widespread application of machine learning to specialized consumer data sources is deliberately designed to extract personal and actionable insights about both individual users and collective behavior.

It would thus be wrong to formulate privacy policy based only on the amount or apparent source of data. One must evaluate the sensitivity of the data as well as anticipate how private or intimate the inferences and predictions that could be made from the data might be. This challenge argues for a privacy framework that comprehensively covers the diverse range of data being used commercially and applies consistent technology-neutral privacy requirements.

Thank you again for the opportunity to testify before you. Machine learning and AI present significant challenges for policymakers because of the rapidly evolving nature of the technology as well as its pervasive use among consumer-facing tech companies in predicting consumer preferences and drawing inferences about their lives. While policymakers should be mindful that machine learning and AI also produce many of the sizeable benefits inherent in consumers' online experiences, such technology enables companies also to both model and shape user behavior. Thank you.

[The prepared statement of Dr. Kearns follows:]

Testimony of Michael Kearns
Professor, University of Pennsylvania
Before the Subcommittees on Communications and Technology and Digital Commerce and Consumer
Protection Of the Committee on Energy and Commerce
“Algorithms: How Companies’ Decisions About Data and Content Impact Consumers”
November 29, 2017

Introduction

Chairmen Blackburn and Latta, Ranking Members Doyle and Schakowsky, and other distinguished Members of the Subcommittees, thank you for the opportunity to testify at this important hearing. My name is Michael Kearns, and I am a Computer and Information Science Professor at the University of Pennsylvania. I am appearing in a personal capacity today, and the views I express are my own. I have been an active and leading researcher and educator in the field of machine learning since the late 1980s, and, in addition to my academic work, I have consulted extensively on the use of machine learning in the technology and finance industries.

The fields of machine learning and artificial intelligence now play a central role in virtually every domain of science, technology, and business in which large data sets and challenging prediction problems are present. The number of instances in which the use of machine learning has provided tangible societal benefits, such as in medical diagnosis and, more recently, agriculture, is large and growing. Machine learning is also used in consumer-friendly activities such as detecting fraudulent banking or credit card activity.

Machine learning also increasingly plays a central role in the data collection and use practices of consumer-facing technology companies. Today I will discuss “data intimacy,” the notion that machine learning enables companies to routinely draw predictions and inferences about consumers that go far deeper than the face-value of data collected as part of consumers’ online activities.

It is not simply a question of whether consumer-facing technology companies are collecting large volumes of data; such companies are collecting information that provides, or allow inferences regarding, intimate details about our personal lives. Search engine queries permit inferences about our financial, physical, and psychological conditions. Social media users routinely reveal opinions, beliefs, or affiliations that might carry social stigma, and that they would be more reluctant to reveal in everyday life.

For example, a recent study showed that, using machine learning, anonymous social relationship data permits accurate identification of romantic partners for over 55% of users --- orders of magnitude higher than random guessing.¹ Another study concluded that Facebook’s data, algorithms, and models are capable of identifying social relationships of which its users are themselves unaware.² It has also

¹ Romantic Partnerships and the Dispersion of Social Ties: A Network Analysis of Relationship Status on Facebook, L. Backstrom and J. Kleinberg, Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing, 2014, available at: <https://arxiv.org/pdf/1310.6753v1.pdf>. See also, Facebook Inches Closer to Finding the Formula for Love, *Wired Magazine*, November 12, 2014, available at: <https://www.wired.com/2013/11/can-facebook-really-predict-our-love-lives/>.

² Facebook Figured Out My Family Secrets, and it Won’t Tell Me How, *Gizmodo*, August 25, 2017, available at:

been established that religious and political beliefs can be accurately predicted from apparently unrelated social, search, and shopping activity undertaken by consumers.

The volume, diversity, intimacy, and modeling of consumers' information has substantial, and rapidly evolving, consequences for consumer privacy and implications for public policy. I will now provide an overview of how machine learning works, the value of information derived through machine learning, and how machine learning is utilized to predict consumer preferences and behaviors.

Machine Learning

Machine learning is the modern science underlying the construction of large-scale predictive models from massive data sets. It is a mixture of topics from areas as diverse as statistics, probability theory, pattern recognition, algorithms, artificial intelligence, and, most recently, distributed systems.

While its origins lie in the 1980s, in recent years, the data explosion enabled by the Internet has made machine learning one of the most important scientific fields, and one that has even entered the popular consciousness. The original efforts to catalog consumers' use of the Internet through hand-coded human expertise or knowledge were quickly overwhelmed by the exponential growth of consumers' online activities. As a result, machine learning has been employed to sift through vast volumes of data to improve the algorithms used for search results and other Internet-related queries.

The algorithms of machine learning and the models they produce are largely automated once in operation. But the development of these algorithms, their improvement and evolution, their implementation in a distributed, cloud-based computing environment, and their specialization to the idiosyncrasies of new and ever-changing data sets remains a highly technical, research-intensive, and human-centric activity. Machine learning has enabled technology companies to create highly predictive models for collective and individual consumer behavior, and to make subtle and accurate inferences about consumers' interests and preferences.

Machine Learning Process

The first step in the machine learning process is known as "feature extraction" or "feature engineering," which are the terms used to describe processes that transform the raw data streams into higher-level abstractions that have more structure, and encode more directly the underlying meaning and intent in the data. Examples include identifying objects and edges in images, or parsing an English sentence in a social media post. The development of algorithms for such feature extractions is actually extremely challenging, and has been the source of many decades of intense research.

Feature engineering turns the raw, unstructured data streams into structured objects with more meaningful and informative representations that are also much more amenable to machine understanding. For many machine learning tasks, the next step is to annotate such data with user feedback, which in the field's terminology is sometimes referred to as "labels" or "supervision." The basic idea is that if individual data items or events (such as sentences, photos, documents, or web pages) can be identified as relevant or irrelevant, good or bad, etc., then one can use sample data to train a predictive statistical model.

<http://gizmodo.com/facebook-figured-out-my-family-secrets-and-it-wont-tel-1797696163>.

The combination of feature extraction with user feedback or supervision sets up a classical statistical modeling problem: the raw user streams have now been transformed into $\langle x, y \rangle$ pairs, where x is some structured representation of complex data items like documents, sentences or images, and y is a signal indicating whether x is “good,” “bad,” or in between. The challenge is then to take a (very) large sample of such data pairs, and build a predictive model – i.e. a model that, given a new, previously unseen x , can accurately predict the associated feedback y . This challenge is precisely the domain of modern machine learning. An example of the end-product of the machine learning process would be a model that takes as input all of a user’s activity on a social network, search engine, or shopping service, and outputs predictions of the ads in which the user would be most interested.

Value of Data Volume, Diversity and Intimacy

Machine learning and artificial intelligence would not be effective without large data sets to analyze. Consumer-facing technology companies in the United States have amassed an almost unimaginable set of data about consumers, both collectively and individually, which enable machine learning and artificial intelligence to draw conclusions about consumer behavior and preferences.

As mentioned above, these large and diverse data sets are the foundation for effective algorithms. Companies compete vigorously to acquire these diverse data sets to support their machine learning capabilities through the development of services as well as acquisitions. For example, search engines provide vast amounts of data about consumers’ interests and the manner in which they conduct searches. Search data can reveal consumers’ financial, medical, and mental conditions. Similarly, mobile operating system data provides a treasure trove of information regarding virtually everything a consumer does on a mobile device.

Services that consolidate location data also provide companies with vast information about consumers’ physical location, and enable such companies to develop inferences about consumers’ activities based upon those locations. Such location information goes far beyond what the GPS receiver in a consumer’s mobile device may divulge because they are amassed using WiFi, Bluetooth, and other technologies as well. Some services seek not only to determine where users are, but where they will be or plan to be in the future; examples include calendar apps, flight and travel shopping services, and navigation apps.

Social media platforms also provide substantial amounts of raw data. In addition to knowing with whom a consumer affiliates directly, social media platforms are able to accumulate information about who a consumer follows or what he or she likes. However, while the quantity of data is critical to develop accurate algorithms, the quality (and intimacy) of such data is important to discern consumer preferences and behaviors.

Use of Machine Learning to Predict Consumer Preferences and Behaviors

Increasingly, machine learning-based algorithms are utilized not only to determine consumer purchasing habits, but also to determine consumers’ emotions. While these algorithms are employed most commonly (and pervasively) to target advertising, as we’ve seen in media recently, such algorithms are also being utilized to generate (or incite) certain emotional responses.

For example, beginning with a substantial set of raw data, researchers recently used a sophisticated “dimensionality reduction” method known as singular value decomposition to

automatically extract a much smaller set of informative “features” to represent each consumer.³ These feature representations were in turn given to a standard statistical algorithm to produce predictive models for each of the targeted categories (sexual orientation, race, political party, etc.). Thus, the raw data on the collective population is transformed into a higher-level model that permits accurate (and intrusive) inferences about specific individuals that were not present in their raw data at all. This model, and the highly detailed information produced by the model, is developed almost entirely through machine learning methods. Other recent research has demonstrated the extent to which people use search engines to express their most private and intimate thoughts and concerns, as though they were entirely unobserved.⁴ When combined with other data sources and the use of machine learning, the detailed insights and predictions that are possible are effectively unlimited.

Public Policy Implications of Machine Learning

From a privacy perspective, perhaps the most important overarching conclusion is that the “intimacy” of consumer data cannot be measured by the number of bits crossing a pipe, or similarly crude metrics that fail to account for the nature, diversity, and content of the data and its potential uses for modeling and inference. It is both possible and common that the highest volume data sources (such as the fragmented and possibly encrypted packets passing through a core router in the Internet) can reveal virtually nothing about the consumers who generate that traffic, whereas much lower-volume and more-specialized data sources can both directly and indirectly reveal the most private and personal details about consumers. In fact, the widespread application of machine learning to specialized consumer data sources is deliberately designed to extract personal and actionable insights about both individual users and collective behaviors.

Thus, it would be wrong to formulate privacy policy or metrics based only on the amount or apparent source of data --- one must evaluate the sensitivity of the data as well as anticipate how private or intimate the *inferences* that could be made from the data might be. And such anticipation, for policymakers or computer scientists, is extremely challenging.

This challenge argues for a privacy framework that comprehensively covers the diverse range of data being used commercially, and applies consistent privacy requirements. Policymakers should also take a forward-looking approach to privacy, and not overly focus on specific data types or practices (which are likely to become obsolete shortly due to the rapidly changing nature of technology). A technology-neutral approach can adapt quickly to new technical and market developments.

Conclusion

Thank you again for the opportunity testify before you today. Machine learning and artificial intelligence present significant challenges for policymakers because of the rapidly evolving nature of the

³ Private Traits and Attributes are Predictable from Digital Records of Human Behavior, M. Kosinski, D. Stillwell, and T. Graepel, *Proceedings of the National Academy of Sciences*, 110(15), 2013, available at: <http://www.pnas.org/content/110/15/5802.full.pdf>.

⁴ *See e.g.*, Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are, Harper Collins, 2017; Essays Using Google Data, Seth Stevens-Davidowitz, Doctoral Thesis, Harvard University 2013, available at: https://dash.harvard.edu/bitstream/handle/1/10984881/StephensDavidowitz_gsas.harvard_0084L_11016.pdf?sequence=1.%25C2%25A0; links to published research articles and items published in the *New York Times*, available at: <http://sethsd.com/>.

technology, as well as its pervasive use among consumer-facing technology companies to predict consumer preferences and draw inferences about intimate aspects of consumers lives. While policymakers should be mindful that machine learning and artificial intelligence also produce many of the sizeable benefits inherent in consumers' online experiences, such technology enables companies to shape commerce, and even belief and emotions. This hearing is therefore an important opportunity for the Members of these Subcommittees to understand and evaluate the risks inherent in such technology.

Mr. LATA. Thank you very much for your testimony. We really appreciate it. And this ends that portion of our hearing this morning. We will now be going to the questions from the Members, and I will begin the questioning and recognize myself for 5 minutes. And again I apologize for my 4 weeks of allergies, and I hope I get better in the next 4 weeks.

Professor Kearns, if I could start with you. Algorithms are used to produce the results that we see on the internet such as when we do a search or see an advertisement. As policymakers, what are the key benefits and risks for consumers associated with these algorithms that we should be focused on as legislators?

Dr. KEARNS. Well, I think the benefits are, you know, pretty obvious to anyone who is a regular user of modern internet technology. The personalization in social media sites, in search engines, and in many other aspects and apps that we use, we all enjoy the benefits of that. I think to me, I think the greatest risks are the kinds of things I talked about, which is, you know, there is sort of a distinction about facts about you and things that can be inferred about you from those facts.

And so it is one thing to, for instance, ask about disclosure or discuss what is actually, literally, in the data that is being collected, but that is kind of where the game is being played, as far as I am concerned. The use of machine learning allows one to make many inferences that are statistically quite accurate about consumers that aren't written down anywhere in the data about that consumer.

So, you know, to give a personal example, the fact that I am an academic and, you know, use a Mac and drive a Subaru probably lets you guess my political affiliation quite accurately already, and if you knew a bunch of other facts about my online behavior, you could probably infer a great deal more. And there are many, many studies these days that sort of establish that fact, and this is a valuable thing to technology companies to be able to do that, to do this kind of—I think in one of the other testimonies here—this kind of microsegmentation.

And I think this is the kind of thing that is hard for people to understand, and it is even hard for the scientists at these companies to understand the sort of power of this, this sort of predictive power that they have. You know, when these models are built they don't really know a priori and maybe even afterwards exactly what properties of consumers or inferences they are making about them that aren't—you know, they go well beyond the latent data itself.

Mr. LATA. Thank you.

Dr. Tucker, your research shows the tension between how much we say we value privacy and in reality how much data we are willing to share online to connect with friends or get personalized recommendations and coupons. What accounts for that disconnect, and how important is the context in what consumers are willing to share online?

Dr. TUCKER. Well, I am really thrilled to be able to talk a little bit about this because I didn't get to mention it in my testimony. And this is a so-called privacy paradox that so many people say they care about privacy but then act in ways which doesn't sort of live up to that.

And one thing, we did a little study at MIT where we showed that undergraduates were willing to share really very personal data in exchange for a slice of cheese pizza. And that was even the ones—and what was slightly disconcerting about it was even the people who said that they really cared about privacy, they usually behave in accordance with those norms, but the moment they saw the cheese pizza was the moment they are willing to share the most personal information.

Now I wish I could tell you that I found any group of consumers out there who were not—or any group of undergraduates who were not willing to share data for cheese pizza, but I didn't. So as of yet, answering your question is hard just because we do see this inconsistency between the way that consumers say they talk about their privacy and actually act out there in the online world.

Mr. LATTI. Thank you.

Professor Ben-Shahar, your research indicates that consumers often view privacy policies as confusing and often ignore them, especially from your photograph. At the same time, mandated disclosure has been embraced in many laws and by many regulators. How should we balance the desire for transparency with the results of your research?

Dr. BEN-SHAHAR. I think we should recognize that our desire for transparency, while well-intentioned and makes sense—very alluring, consistent with all American ideologies—all these transparency laws and mandated disclosure laws pass without opposition in this chambers or in any State chambers. This is the one unifying American law. I think we should also recognize that there is a good reason probably why it is so easy to enact these laws: There is nothing to them.

And therefore I think that it is important to set them, cast them aside, and then that would enable us to actually get into the—I think in my book I give the example of medicine in the 19th century. Almost every disease was addressed by blood-letting. It took the ability or, you know, from the medical profession to recognize that this is, you know, that panaceas don't work. You cannot use that to start figuring out solutions for each individual problem.

And today you are talking, you know, I am invited to talk to you about data policy. I was invited by the FTC and before other agencies to talk about consumer lending, other contexts in which transparency and disclosure is the key regulatory technique, and I keep suggesting to them that it is in your area. You have to first ask yourself what the problem is.

I think it is striking to hear what Dr. Tucker and others are finding, that statements about the magnitude of the problems are not matched by the behavior and economic reality. Data privacy is a nice kind of buzzword and data security we are really worried about, we can brandish the number of people that were hurt by the different—were implicated by the different breaches that occurred, security breaches.

But what is the evidence about actual consumer harm? Most of the lawsuits that followed, you know, the lawsuits that have followed the Target breach and the Equifax breach were by merchants, credit card companies, banks, they are suffering a lot of the—because our laws largely protect consumers from these inci-

dents. So I think I do not want to suggest that there is no harm in these areas, but it is critically important to understand its magnitude before we begin to think about solutions.

Mr. LATTI. Thank you very much. And, since I ran over, I will recognize the gentlelady from Illinois, the ranking member of the subcommittee, and also give you a little more time on your questions.

Ms. SCHAKOWSKY. Thank you. You know, it is hard to decide who to really focus on because we only have 5 minutes. You know, when it comes to transparency, not only don't I take the time to read it, but in order to get to my goal if I don't hit Accept, I Agree, then I can't finish the transaction. So most of the time, for both reasons, I just accept and move on.

But I do want to talk about enforcement, and therefore I want to ask Ms. Moy some questions. In Chairman Blackburn's opening statement she talked about shifting privacy from the FCC, the Federal Communications Commission, to the Federal Trade Commission, so I think it is important to understand how the FCC and FTC differ, you alluded to that. But so, Ms. Moy, can you briefly describe the FTC's authority, if any, to issue regulations?

Ms. MOY. The FCC or—I am sorry, the FTC really doesn't have authority to issue regulations. It can issue rules under—it can issue Mag-Moss rules, but it is extremely difficult to do that, and as a practical matter nearly impossible. It can issue rules under the Children's Online Privacy Protection Act and has done that rather effectively, and the Safeguards Rule under GLBA.

But when it comes to general privacy and data security obligations, the FTC is unable to issue regulations.

Ms. SCHAKOWSKY. So the FTC can't use the typical notice and comment rulemaking process to issue regulations about what personal information platforms can collect from users or how those platforms can use that personal information to determine what content it shows to users, correct?

Ms. MOY. That is right.

Ms. SCHAKOWSKY. So which means the Commission is limited to bringing enforcement actions after unfair, deceptive practices have been committed, and often after consumers have been harmed already, right?

Ms. MOY. Yes.

Ms. SCHAKOWSKY. So let's talk about the FTC enforcement tools. In your written testimony you wrote that, quote, "the FTC generally can only take enforcement action against entities that use consumer information in ways that violate their own consumer-facing commitments." Can you describe what do you mean exactly by consumer-facing commitments, and are you referring to policies like the terms of services and privacy policies?

Ms. MOY. That is right. The bulk of the FTC's privacy and data security authority comes from Section 5 of the Federal Trade Commission Act which authorizes it to prohibit unfair and deceptive trade practices. As a practical matter, the FTC almost never enforces unless it determines that there is deception that has occurred, and it evaluates a possible deception based on something that a company has said perhaps in a privacy policy and then trying to figure out whether or not it has violated that.

Ms. SCHAKOWSKY. Even when a platform does violate its own policies, the FTC's remedies are limited. As you noted in your written testimony, the FTC cannot impose a fine against that platform. What are the remedies available to the FTC?

Ms. MOY. Exactly. Yes, you know, and as I mention in my comments, I think the authority of an agency is only as good as its enforcement is. And when it comes to the FTC, although it can bring actions for deception when as it relates to privacy and data security, with few exceptions it cannot levy civil penalties against companies that violate privacy and data security commitments. And as a result there is very little in way of teeth when it comes to the FTC's authority.

Ms. SCHAKOWSKY. So I know that both Acting Chairman Ohlhausen and Commissioner McSweeney support giving the FTC civil penalties authority, and I believe you do, too, as well. Is that right?

Ms. MOY. That is right.

Ms. SCHAKOWSKY. And do you think it would benefit consumers if the FTC had authority then to issue regulations under the normal notice and comment process?

Ms. MOY. I do. I think that the fact that the vast majority of consumers are asking for greater consumer privacy protection and for the law to be stronger in this area suggests that we would benefit greatly from greater authority for the FTC or another agency.

Ms. SCHAKOWSKY. Well, so are there other things that Congress can do? I mean, you alluded maybe to other agencies to help strengthen the FTC's ability or some other agency to protect consumers.

Ms. MOY. Well, in addition, as of right now the Federal Trade Commission can't actually regulate the actions of common carriers, and that is a major problem that, particularly with a recent case or a case that is currently pending in the Ninth Circuit, it is unclear whether the FTC has any authority at all to enforce the privacy and data security obligations and activities of companies that have any common carrier practice at all. So internet service providers that offer—whether broadband is classified under Title II or not, the FTC may well not be able to.

Ms. SCHAKOWSKY. So in the short term what should we be considering?

Ms. MOY. In the short term I think that we do need strong protection, privacy by default, ideally, for entities where consumers have no choice but to share information. And I also think that we need to preserve existing protections. We need to preserve existing protections at State law as well as existing protections under regulations like net neutrality.

Ms. SCHAKOWSKY. Thank you. I yield back.

Mr. LATTA. Thank you. The gentlelady yields back.

The Chair now recognizes the chairman of the Communications and Technology Subcommittee for 5 minutes.

Mrs. BLACKBURN. Thank you, Mr. Chairman, and thank you all for your testimony. It is so enlightening, and we appreciate it.

And Dr. Kearns, I am going to come to you first. Thanks for the work you are doing on privacy and around those elements, and we have had a lot of focus on privacy here. And earlier this year I had

introduced the BROWSER Act, and basically it has two guiding principles, things that many of us think are very important. One is that we have to find a better balance on privacy moving toward giving the consumer more information and more control over, as I term it, their virtual you, their information that is being collected and used and sometimes distributed; and then, secondly, that consumers have the very same privacy expectations across the entire ecosystem. They are not distinguishing between the ISPs and the edge providers, so when we are setting the ground rules on privacy, they should reflect that.

So I would like to hear what your thoughts are on those two points. And when we are talking about online privacy, do you think that people make that distinction? When we are talking about appropriate balance, where is that appropriate balance within opt-in where the consumer owns that information or either opt-out? So I would love for you to talk about that for a minute.

Dr. KEARNS. Yes. These are good questions, hard questions. First of all, to preface, I don't have specific policy recommendations on these issues. But as a scientist, when I think about the landscape for consumer privacy, the first thing I think about is kind of how actionable the data being collected is and sort of at what level of abstraction it is. And furthermore, there is a phrase I like to use, which is "data triangulation," which refers to the incredible power you can get from having multiple sources of data about the same individual.

So to me, you know, when I think about privacy, the things I worry most about are cases in which there are parties that are collecting sort of very private, intimate data on the one hand and also many different sources of it. So, to give an example, you know, by seeing what you buy I can know a lot about you. By seeing, you know, what you search for I can know much more about you. By knowing not only those things but where you are, that gives me a great deal of more information. And if you, for instance, let me also maintain your calendar for you, then I also know where you will be in the future.

And I think that the, you know, greatest privacy concerns I have are at that level, at the level where people are very directly expressing, you know, things that might be quite private, things that they wouldn't express in public forums, or that they are expressing in a public forum like a social networking service but are completely unaware how strong the correlations are between their own behaviors and their friends' behaviors and their other online behavior.

And so I think in terms of helping consumers understand the privacy landscape it is important not to ignore any source of data. I am not claiming that ISPs, for instance, aren't also collecting very large amounts of data, but to me, I personally am much more concerned about the data I kind of willingly give away using a search engine, and then also letting my operating system track my location online or the presence of beacons in retail stores that kind of correlate my online and my offline behavior.

Mrs. BLACKBURN. Great.

Ms. Klonick, I want to come to you on economic incentives and the economic incentives that the platforms have to use algorithms

to curate selective content. And I think Dr. Kearns used the term “microsegmentation” as they are looking at that for users, you know, based on this online activity. Would you agree that the platforms are being paid to prioritize certain content over other content? And touch on the free speech implications there.

Ms. KLONICK. Yes. Inasmuch as advertised content is paid content over their user content, I think that these, you can absolutely prioritize certain types of content. I am not familiar with the algorithmic processes that would prioritize one user’s content over the content of another and that they are being paid to do so right now, but the free speech implications of the vast power of these platforms to self-regulate is, they are twofold.

One, it has a lot of implications for the user’s speech rights in how these private platforms can unilaterally control at what goes up and what stays or goes down on their sites. But also these platforms have free speech rights, arguably, free speech rights, themselves. So their right to create the community at Facebook or at Twitter, for example, is arguably their own First Amendment right.

Mrs. BLACKBURN. Dr. Tucker, on the economic incentives, do you think that some of these platforms should be willing to pay consumers or users more than a free slice of cheese pizza?

Dr. TUCKER. Wonderful question. Now, this is a very interesting question. So the slice-of-cheese-pizza example was really about the consistency between what people say about their privacy and then how they act.

Now, in terms of paying for data, there have been many experiments, some of them launched from Cambridge, Massachusetts, where various startups have helped devise, have tried to actually set up markets for data. And the reason that is so attractive is, from an economics point of view, one way of thinking about privacy is, really, there is a lack of clarity about property rights. So a market for data is an attractive notion.

Now, in all of the instances, though I have been really excited at the beginning because of the idea of actually setting up a market for data and paying consumers, all of these platforms have failed for the simple fact that the kind of consumers they attract who want to exchange their data in these markets tend to be, how can I say it, the less commercially exciting consumers. And we have had this problem of actually just setting it up, making these markets work just because we haven’t been able to get the right set of consumers.

So I think it is a wonderful idea. I hope one day we will get it to work. We haven’t yet.

Mrs. BLACKBURN. Yield back.

Mr. LATTA. Thank you very much. The Chair recognizes the gentleman from Pennsylvania, the ranking member on C&T, and I also yield you the long time, too.

Mr. DOYLE. Well, Mr. Chairman, this terrible precedent that you have started by allowing everybody to go 2 minutes over, I am going to try to get us back on track and just use my 5 minutes.

You know, when you think about all this technology—social media, the internet, artificial intelligence—you know, the most wonderful, horrible invention in the world, and as consumers we tend to look at the bright side of all this technology without under-

standing the dark side. If anybody thinks they have privacy, the only way you have privacy today is, I call it to go Flintstone, to have the old flip phone, to not be on Facebook or Twitter or any of these social media sites.

But, you know, the reality is, for most Americans over 80 percent of the land mass of country, most Americans have only one ISP provider. They don't even have choice when it comes to that. And so they go on their ISP, and it is the only one they have, and they tell you how they are going to use your data, and it is about 20 pages long of a bunch of legal jargon that most attorneys probably couldn't understand. And if you don't click I Agree, that is it, you don't have access to any of this.

So you don't even need a cheese pizza to get people to give up their information. They want to go online to do whatever it is they want to do online, and the only way they can get there, especially if they only have one ISP, is to do that. Now, search engines, you have some choice and you can read different, you know, policies on search engines of how they use your data, and it varies online, whether you are on Google or whether you are on DuckDuckGo or these various sites, at least you have some choice. With your ISP, most Americans don't have choice. They have one place to go.

And it is kind of ironic that we are here today to discuss concerns about algorithms used by these social media companies to curate content on the internet, but as we speak, over at the FCC the Chairman is getting ready to allow broadband providers to block and edit speech on the internet at their discretion, relying on public commitment from these providers that they are going to behave.

And, given the Ninth Circuit case casting doubt on whether the FTC may even police these broadband companies, it is sort of creating a situation where broadband companies are just free to reign over consumers with impunity, and the FTC for all intents and purposes is a toothless tiger. We talk about shifting all this watchdog function over to the FCC—

Ms. SCHAKOWSKY. FTC.

Mr. DOYLE [continuing]. And they don't really have the ability to do anything on behalf of consumers. Right now, if this law passes on net neutrality next month, basically there is no law of the land, we are just trusting people to behave. They are saying they are going to behave, and we are going to take them at their word that they are going to behave.

Professor Moy, I wonder if you can give us some examples of how broadband providers behaved prior to the enactment of strong bright line rules that were put in place by the FCC in 2015?

Ms. MOY. Thank you, Representative. That is a great question. Right, because before we had rules we did see broadband providers, internet service providers, blocking things like Voice over IP, blocking tethering applications, so they could extract more from consumers in monthly fees, blocking peer-to-peer sharing applications. AT&T threatened, I think, to block FaceTime unless consumers agreed to pay more for the ability to use that.

So, you know, we certainly have seen examples in the past of ISPs using their power as gatekeepers to prevent consumers from using services that may well want to—

Mr. DOYLE. So tell me what recourse would consumers have if the FCC Chairman gets his way and removes these protections?

Ms. MOY. It is hard to see how they would have any recourse at all. I mean the FCC plans to rely on the consumer-facing commitments again of ISPs, but it is unclear whether ISPs would actually be required to commit to not prioritizing content, not blocking content. And even if they did make those commitments and then violated them, the FTC—you know, you mentioned the Ninth Circuit case—may not be able to enforce against them. You know, their enforcement authority against ISPs is going to be questionable at best or nonexistent at worst. And even if they could enforce, again, they don't have civil penalty authority.

Mr. DOYLE. Thank you.

Mr. Chairman, in the spirit of staying within 5 minutes, I have 5 seconds left, and I will yield them back to you.

Mr. LATTI. Thank you very much. The gentleman yields back. And at this time the Chair recognizes the gentleman of the full committee for 5 minutes. The chairman, the chairman of the full committee.

Mr. WALDEN. OK, thank you all, I appreciate it. And thanks for our witnesses. My apologies for having to come and go a bit today, but we do appreciate your written testimony and the answers to our committees' questions.

I guess, Dr. Moy, the question I have because we are concerned about misbehavior by ISPs, I am also concerned about misbehavior by others in the ecosystem of the internet. And it strikes me that on these information platforms we have seen foreign actors try to affect our elections with paid advertisement that is targeted.

We know that there is, in effect, paid prioritization on some of these platforms, right, because you buy advertising, and it strikes me that at least Google—it is an amazing American company, it does incredible work but has about 77 percent market share of search, and I have had consumers complain to me about what they believe to be the use of algorithms that have disproportionately affected them.

So what—and maybe this can go to everybody on the panel, but so if, who governs the edge providers when there are questions about use of private data or—nothing is private anymore, but your data and how that gets—and I don't mean this in a negative way, but manipulated use through the algorithms, which we are all trying to get a better handle on, so who governs their activities and what enforcement protocols are in place for those?

And I will start with you, Ms. Moy.

Ms. MOY. Great. Thank you so much for the question. So yes, right now those practices are, in theory, governed or regulated by the Federal Trade Commission, enforced by the Federal Trade Commission, again under this idea that they can enforce consumer-facing commitments. But, you know, I think you raise a really good point, which is that the growing power of these platforms to editorialize on content is potentially problematic, and we should explore possible solutions to that.

But in the meantime, the last thing that we should be doing is eliminating protections that consumers have against paid

prioritization at the network level, where there is very little transparency.

Mr. WALDEN. Right. But in terms of other enforcement in the overall ecosystem, if I have a complaint against a search engine or I have a complaint against my social media, I go to the—my only recourse is the Federal Trade Commission, which you have said doesn't have the kind of enforcement authority you would like to see it have, correct?

Ms. MOY. Right, right. Yes, indeed. And, you know, and staff and Commissioners—

Mr. WALDEN. Do you think there should be greater authority for enforcement over the edge providers or similar to what you would see over the ISPs?

Ms. MOY. I would certainly support adding protections for consumers across the board. I think that there are concerning practices by both types of actors. I would caution this committee against exploring a one-size-fits-all solution to—

Mr. WALDEN. Why?

Ms. MOY. Because I think that, you know, again the types of information that various actors have access to is different. The commitments and relationships with consumers that they have is different. For example, consumers are paying dearly for monthly access to the internet with a broadband provider, whereas they often are getting certain other services for free or—

Mr. WALDEN. Right. No, it is an exchange of value. Yes.

Ms. MOY. There are certainly differences between different types of actors as well the availability or lack thereof of sharing information with a particular provider or particular type of actor.

Mr. WALDEN. So let me ask you a question, because we have also heard before this committee that there is a very high rate of encrypted data that passes through the ISP pipes, if you will allow me to use that term, and that that is encrypted. They don't know what those data are. It is encrypted, it goes through. It is well over 50 percent, perhaps, so they don't see it, but the other platforms do see the data and can use it and do use it in that exchange, as we know. I am not saying these are bad things.

And I think we have heard—I believe it is Dr. Tucker. I am going to get them to make those nameplates bigger for us old people that have vision issues. But the point is that they can, they see it differently. Can you address that, the differences you have seen in Europe versus here maybe on how our technology has expanded dramatically and innovation here because we haven't cranked down as much, right, on privacy?

Dr. TUCKER. OK. So in the past—and this was about 2011—I did research on how some of the early European data privacy regulation really stymied the ability of Europe's ability to create additional ecosystem like we have now. And since then there has actually been follow-up research which has shown that it wasn't just at the beginning, but it has kept on going, and we have seen an awful lot of lack of entrepreneurship in Europe, too.

And so we have seen the failure at the beginning and then the follow-on failure of entrepreneurship, and I think to me that is what has really distinguished what we have seen in the U.S. tech sector.

Mr. WALDEN. So we have had, am I accurate to say we have had more of a light touch regulatory approach to the internet up through 2015 from Europe?

Dr. TUCKER. I think it is certainly true that we have had a sector-specific touch, right. That we have focused on areas we might care about such as health, private financial data, children, rather than going for a broad brush approach.

Mr. WALDEN. All right. I have exceeded my time. Thank you all again for your testimony, it is very helpful in our discussions, and I yield back.

Mr. LATTA. Thank you very much. The chairman yields back his time. And at this time the gentlelady from California, Ms. Matsui, is recognized for 5 minutes.

Ms. MATSUI. Thank you very much, Mr. Chairman. I want to thank the witnesses for being here with us today.

I have a question, I think, for Ms. Moy right now. In 2015, the Office of Management and Budget issued a memorandum requiring all publicly accessible Federal websites to only provide service through an HTTPS connection by the end of 2016, which was last year. HTTPS protocol ensures that a consumer's connection is encrypted from their devices all the way to the Federal Government's systems. Regular HTTP connections sent in plain text can be intercepted and exploited by anybody or anything between the user and the website, including somebody using public Wi-Fi. A study released earlier this month revealed that only around 70 percent of Federal websites employed HTTPS protocol.

Ms. Moy, how important are the security standards like HTTPS to protect the confidentiality of internet-delivered data on both Federal and commercial websites?

Ms. MOY. HTTPS is very important. HTTPS would encrypt in transit the information that is transmitted via websites. So, for example, if you fill out a web form, for example, perhaps in an application for a service that you might find on a Government website, and that form contains or asks questions about information that is highly private, such as information about financial status or personally identifying characteristics like Social Security number, then, if the site is not employing HTTPS technology, one could mount an attack on the transmission and potentially read the information that was transmitted.

Ms. MATSUI. So how would you know whether it employs the HTTPS on the Federal website?

Ms. MOY. So this is the type of thing where in a browser bar, you know, you will see up at the top the little, now we have that little icon, the little green lock that indicates trust for HTTPS protocol.

Ms. MATSUI. OK, something what we never look for, anyway. OK, thank you.

I want to talk about embedded networks. Across almost every industry, we are seeing a trend towards embedding communications functions into their structures. Applied data science such as a massive internet of medical things, rely on faster, more efficient, and more robust communications with innovative enabling technologies such as blockchain. Blockchain can facilitate the exchange of massive amounts of data, but as a decentralized ledger technology it

can make online transactions faster and cheaper while maintaining and protecting data integrity.

Anyone on the panel, how can new digital technologies and applications help consumers improve data security? Anyone want to start on that one?

Dr. TUCKER. Well, I have written a little bit on blockchain, so I am just so excited that you mentioned it, and I am glad that you mentioned it without mentioning bitcoin, which is always a distraction.

Ms. MATSUI. It is a distraction.

Dr. TUCKER. And certainly we have got an initiative at MIT which gives enormous optimism for the kind of process that you are describing where, really, what we call verification costs for making these kind of transactions easier.

Do I have any caveats? My only caveats are that when we have studied it, and if we are thinking about blockchain as being a recipe for protecting privacy, that in some sense it can sometimes embolden people to be somewhat more careless about their data surrounding the edge providers who are trying to serve the blockchain. And so, for example, we have seen that the mere mention of blockchain encourages people to share really quite personal information such as telephone numbers and so on without any guarantees of protection.

Ms. MATSUI. So they feel like it is much more safe because of the blockchain. They just figure that what they have heard about it, that this is a safe way to go?

Dr. TUCKER. Yes. That is right. So I sort of have the analogy that it is a bit like, once you have your seatbelt on, perhaps you drive a bit too fast, that kind of an analogy. And so I think it is definitely a step forward, but we have to realize that of course it is going to interact with other providers, and the most will be privacy concerns there.

Ms. MATSUI. Thank you.

Did you want to make a comment?

Mr. PASQUALE. I would just say very briefly that I testified in September before the Senate Banking Committee, and I mentioned in part of my testimony futurist financial technologies such as blockchain. And I think that it is just very important to distinguish between the private permission blockchain and the public permissionless. I have a lot more confidence in the sort of private permission because it involves what I call complementary automation technology complementing individuals rather than replacing them.

So I think that it is, just in terms where I have hope, it is more in that latter category of private permission blockchain.

Ms. MATSUI. OK, thank you. And I see my time is expired. I yield back.

Mrs. BLACKBURN [presiding]. The gentlelady yields back. Ms. Matsui, I just mentioned to counsel that we may want to secure his Senate testimony and submit that into the record in coordination with your question.

Ms. MATSUI. Thank you very much.

Mrs. BLACKBURN. Agreement? So ordered.

[The information appears at the conclusion of the hearing.]

Mrs. BLACKBURN. Mr. Shimkus, 5 minutes.

Mr. SHIMKUS. Thank you, Madam Chairman. It is great to be here. I got to listen to your opening statements. I found them all very interesting. And then I had to run upstairs to do Energy Markets and Interconnectivity, and then I came back down here, so I may have missed a few issues.

I just want to put on the record on this whole net neutrality debate, it is just, for a lot of us it is what is the enshrined law by the legislative process versus what a regulator decides what to do. And what we are seeing now with the passing of the Obama administration, and the Trump administration, is I kind of explain to my constituents it is a pendulum. We are going to do it this way, now we are going to do it this way, now we are going to do it this way, and to stop the pendulum you have to pass a law. You have to enshrine that into a statute, and I would encourage my colleagues to come together to do that.

I also want to incentivize build-out. I like more pipes versus less pipes, and I don't want the Government deciding how one pipe should be structured. I would rather have so many pipes that everybody gets what they want when they want it at the speed that want it, and if you are a market-based conservative you have got to send a price signal.

And then the other issue on that is this whole—part of this was kind of paid prioritization, or we are talking about so small of lag of time that I can't even use the proper terminology. But would I rather have lifesaving telemedicine go fast versus a Three Stooges video? The answer is yes, I would. So I just want to put that in the guise of some of the debates based upon what the FCC is considering. And then I want to segue real quick to this whole—this is a fascinating panel because you all have done, brought pretty much a different focus and sometimes there are similarities on privacy, on algorithms, on data.

So I want to use this example. Over the Thanksgiving break I visited Washington University, a major medical facility in St. Louis, and so I briefly drew my little DNA strand, right, here. And so the question with data is in the healthcare arena we want to go to drive to personalized data, I mean personalized medicine, and personalized medicine means we understand the DNA sequence, and we can pull that out. So then a cancer patient, we don't have to try 15 different types of cures, we can direct it.

Now that creates a lot of issues public policy-wise. One issue is the data collection. The other one is data sharing. The other issue is privacy. And when you are doing medical research, I mean, you are really trying to share that data, that DNA sequence of this one case across different major schools of medicine across the country and probably across the globe.

So that goes to a lot of your individual comments. I kind of want this to happen. I really believe in personalized medicine. I think it is going to be a huge savings, and I think it helps treat the patient quicker and return them to a very, you know, return life. And we have these hurdles that we are all discussing here.

Anyone want to weigh in on—Mr. Pasquale, and then I will go to Dr. Tucker. I got about a minute, 2 minutes left.

Mr. PASQUALE. I will be very quick to say that I completely agree with you, and I think that, you know, we have talked to—I run a health law podcast with Nick Terry called “The Week in Health Law,” and we talked to several people who are law and policy experts in this type of area, sensitive health data, and we get a lot of good advice on, you know, how can we develop best practices in order to enable data liquidity, data flow between institutions.

But I would also say, you know, based on some of the great work done by Sharona Hoffman in her article “Big Bad Data,” that sometimes if we don’t have good data practices so we know where data comes from and where it is going to, that may impede the scientific validity of some of the findings. So I think we have heard a lot about privacy impeding innovation, but there are ways in which good data practices, good record keeping, can actually help promote innovation as well and promote scientific validity.

Mr. SHIMKUS. Thank you.

Dr. Tucker?

Dr. TUCKER. So I have a study coming out, it is forthcoming at Management and Science, where we actually look at different types of regulation and how they promote or don’t promote the kind of personalized medicine you are talking about. And what we found there was that basically just focusing on consent was really quite harmful to patients being willing to adopt this kind of or sort of give this kind of unique data in a cancer treatment setting.

What did seem to work, though, was actually giving control to patients, and there were some States that actually experimented with creating ownership or property rights over genetic data, and we have actually seen quite a bit of efficacy in terms of promoting personalized cancer treatments in those States.

Mr. SHIMKUS. Anyone else want to weigh in? I really enjoyed—again I am having a hard time, too, with Mr. Ben-Shahar on the statements of—I mean, how many of us get financial booklets after the fiscal year, and how many people throw it away? I bet you 99.99 percent of all people who get those booklets on what you should know. And I think it is a protection. It is really a protection for those people who are controlling our data. “OK, we have done it. We have given you the information, now it is your fault if you don’t follow it.”

So, it is a great hearing. I appreciate everybody being involved. And I yield back my time.

Mrs. BLACKBURN. The gentleman yields back and, Mr. Green, 5 minutes.

Mr. GREEN. Thank you, Madam Chairman. And I want to thank our two chairs and two ranking members for the hearing today, and as well as our witnesses.

It is pointed out that personalized content that we all see on various online platforms is curated by both humans and algorithmic technology. However, the potential for harm from algorithms can be particularly difficult for Congress to address, and thus we should be focusing on it.

Professor Kearns, in your testimony you point out that machine-learning-based algorithms can be used to determine a consumer’s emotions at any given point in time. How do you monetize that?

Dr. KEARNS. Well, the short answer is I don't know. But certainly, if I can shape people's moods and it seems plausible people might be more willing to shop if they are in a good mood rather than a bad mood, that might be one way that I could monetize it. I think more generally, though, knowing detailed, fine-grained information about people's mental and emotional states in addition to, for instance, knowing about medical facts about them and their fitness level and their financial health, et cetera, I mean, it has clear sources of monetization.

And some of my colleagues on the panel have mentioned some of the negative ones already, such as targeting groups that are particularly vulnerable at a particular time. There is a great deal of documentation, for instance, on kind of predatory loan practices online in the arena of for-profit education, for example.

Mr. GREEN. OK, thank you.

Professor Pasquale, if a person often does online searches for phrases that might signify challenging financial circumstances such as financial counseling, how might that change the ads and the search results that they see online?

Mr. PASQUALE. Oh, that is a terrific question. And one of the big worries that a lot of advocates have is that we can route people into different opportunities. So, for example, if you have exactly the type of searches that you are mentioning, someone might be routed towards payday loans, others might be routed away from them. Now to Google's great credit, I think, 1 or 2 years ago, working actually with Georgetown, they started some self-regulation where they said, "We are not going to have certain ads on that are over 36 percent APR." And I think that is very important, but I also worry that, you know, kind of competition concerns might arise if, for example, Google owned its own finance company that had a business model that would be advantaged by that particular rule.

So I think we have to balance, you know, we have to both encourage tech giants to try to self-regulate to avoid the type of tracking that you are invoking, but we also have to have outside authorities to be able to watch that self-regulation, as well.

Mr. GREEN. Or just so the consumer knows that, you know, that is being done and you might not be getting some other offers, that somebody else is making that decision on what they are presenting to you.

Another question I have, you mentioned in your testimony that in 2016 after Facebook was found to be enabling discriminatory housing ads, it promised to change the system to address that issue but has not done so. Could you talk about efforts that Facebook and who might require Facebook to fix this problem, and why they may not be successful?

Mr. PASQUALE. Yes. I think that the issues here are, it is a complex ad ecosystem and so there are lots of different moving parts in the ads, but I think that what is disappointing is that there was this expose in ProPublica, there was a lot of attention to it. There were pledges to do better, but we just saw in the past week or so that the same people that exposed the original problem, that they are saying it hasn't been solved.

So I think that is, again, another example where we have to empower either State or Federal regulators to actually have some

teeth and to impose some of the penalties that would actually lead to a positive response.

Mr. GREEN. As I found out in this job, everybody needs the boss and has to answer to someone. So we don't have an agency that can do that right now with Facebook if they agree to do something and do not do it?

Mr. PASQUALE. I think that there are possibilities with respect to, say, the deceptiveness or unfairness authority at FTC. I would also have to research with respect to the Department of Housing and Urban Development and its own enforcement practices, but that is not something that I have personally looked into, so I would have to look into that. Yes. And I could send that later on to the committee, yes.

Mr. GREEN. Professor Kearns, you advocate for a policy approach to the extraction of consumer data that is technologically neutral and accounts for the sensitivity of the data collected. My question, can you elaborate on what you think that policy might look like?

Dr. KEARNS. Yes. I mean, first of all, maybe let me take the opportunity to say one thing that I think has been running through my head but I haven't been able to get out yet, which is especially on issues of discriminatory behavior by algorithms, I do think that there are scientific things that can be done to address this and there is a, you know, not small and growing community of AI and machine-learning researchers who are trying to design algorithms explicitly that meet the various fairness promises and guarantees.

And it is still very early days, but this sort of idea of endogenizing some kind of social norm like fairness or privacy inside of an algorithm I think is extremely important, because while regulatory and watchdog agencies will always be very important, you know, the way a computer scientist would put it is they don't scale, right. So, if instances of malfeasance or privacy or fairness violations have to be caught by human organizations looking at, you know, specific instances or behaviors, they just won't keep up, right, because the tech companies are doing this at massive scale in an automated way.

In terms of what can be done, you know, I think it is possible to audit algorithms for various kinds of behaviors without compromising the proprietary nature of the models or algorithms used. And a rough analogy I would offer are kind of the stress tests that banks have been subjected to on Wall Street where, you know, you have to demonstrate certain properties of behavior of your algorithm, but you are not, you know, releasing the source code for it.

And I, you know, without having super-specific suggestions in that regard, I think that that is a promising general direction for policy and one that can balance between, you know, a company's legitimate right to preserve their intellectual property and consumer and societal concerns about the behavior of those algorithms.

Mr. GREEN. Thank you.

Mr. Chairman, I know I am over time. I appreciate your courtesies.

Mr. LANCE [presiding]. Thank you very much, Mr. Green, and I recognize myself.

Ms. Klonick, in your testimony you mentioned choice as a key part of regulators' decisions not to pursue Title II-like regulations for online platforms. Title II-style regulations may be inappropriate for edge providers or for others in the internet ecosystem, as well. However, some have argued there are fewer choices among online platforms because each website or application serves a specific audience with a specific service. Would you please comment on that? Thank you.

Ms. KLONICK. Yes. I agree with that statement generally, that specific platforms speak to a specific audience. But there is an enormous and incredibly important distinction to be made here, and that is that there is a huge difference between companies that have kind of natural monopolies like ISPs and then content platforms like Facebook and Twitter.

And the former kind of a piece of the pipe, or to put it in terms of speech, they are kind of the printing press and you don't want the printing press rearranging letters or blocking out sentences. You want it to be content-neutral to a certain extent, but you do want the paper or the writers or the editors who use that printing press to be able to make decisions based on the content, and that is something why what we are talking about today is so important.

Mr. LANCE. Thank you. If there are fewer choices among these platforms, how does that change the evaluation of the platform's ability to moderate content? Does it make it more or less troublesome in your judgment?

Ms. KLONICK. Yes. I think that as Representative Doyle said earlier, that one of the issues here is that there is a lack of choice between certain types of providers, but on these platforms right now there is just a plethora of choice. I mean, Twitter might have a monopoly over 280 characters of text and Facebook might have a monopoly over a kind of like a relatively safe, family-safe community, but there are plenty of other presences that are currently online. Of course, if that changes in the future and the taxonomy of what these different platforms are able to provide and what users use them for and how they end up having a monopolization or not over broader areas, then I think that that is something that can be revisited.

Mr. LANCE. Thank you very much.

Professor Ben-Shahar, as many of the online platforms we are discussing today offer their services free of charge to consumers, how do we as lawmakers evaluate the appropriate balance between personal privacy against convenience?

Dr. BEN-SHAHAR. Thank you very much for the question. I was hoping to be able to say a few words about that. I think we should be very careful not to change this grand bargain, people paying for excellent services that they like very much not with money but with their data. And it would be a, I think, disaster of consumer protection if we changed that, if you ask consumers in the aftermath of some reform that removed that bargain and made them pay for things like Google, Facebook, and other things with money, if they feel that they were helped, I think they would say in unison, "No, don't do this."

In that sense, I think the bargain and the underlying bargain is an excellent bargain. Now, there are worries that of course arise,

and I think this is the ultimate, the foundational problem of data policy. It is not privacy or security, it is competition. It is the fact that there are very few companies that dominate the central forum in which these exchanges occur—Google, Amazon, Facebook, and maybe a few more small players.

I am not so worried about the ISPs. They, notwithstanding the fact that on broadband there is some local monopolies, there is great competition from mobile, but these big three, or big four if you throw in Apple, big five if you throw in Microsoft, have a lot of power, and the FTC has failed, for example, last year, to intervene in something that the Europeans thought, I think rightly, as raising antitrust concerns.

So to conclude, I think that the concern for consumers will arise from lack of competition and concentration, not from privacy and security.

Mr. LANCE. Thank you very much, and I yield back 42 seconds and I recognize Mr. McNerney of California.

Mr. MCNERNEY. Well, I thank the chairman and I thank the witnesses. Sorry, I missed some of your testimony a little earlier. Professor Moy, what do you think the benefits of the current FCC rules for consumers and small businesses are regarding net neutrality?

Ms. MOY. Great, yes. So I mean that is a great question. I appreciate that question. The current rules enable small businesses to reach consumers. That is the short answer to the question. You know, if we didn't have rules that prevented ISPs from paid prioritization and blocking, then it would be much more difficult, potentially, for small businesses to reach consumers.

Mr. MCNERNEY. So you would agree—or I don't want to put an answer in your mouth—would you agree that it would be harder for small businesses to innovate if the FCC Chairman's proposal is adopted?

Ms. MOY. Yes. You know, and it might even be very difficult for a business to know whether or not it is being throttled if it is being throttled. The draft order has transparency provisions in it, but it is unclear whether the transparency provisions would be consumer-facing or in fact if some companies could fulfill those by just turning over information about their practices directly to the FCC.

Mr. MCNERNEY. Well, that sort of leads, already answered my next question. But the new rules or the new regime would require or ask businesses if they feel like they have been subject to anti-competitive practice to go to the FCC to resolve their problems. How quickly do you think the FCC could respond to those sorts of requests?

Ms. MOY. I mean, if it could respond at all, I mean, well, I think the question is whether it could respond at all, right. So there are many practices that might seem anticompetitive but not raise to the level of an antitrust violation. So, for example, if an ISP were throttling a service that an innovator is introducing into the market but that doesn't compete directly with the ISP service of a phone or internet provision, then that practice might look anti-competitive but might not be considered an antitrust violation.

Also if, you know, if a company were to try to bring an action in court, you know, I think there is this idea that companies might

be able to bring antitrust actions in court, but antitrust actions in court take many years and may cost potentially millions of dollars to mount against a major incumbent. And that can be, you know, that is a barrier that really creates impossibility for a small business or—

Mr. MCNERNEY. Sure. And what sort of penalties could the FTC impose, and would they be effective?

Ms. MOY. Right. I mean, so again the FTC's primary authority when it comes to enforcing something like net neutrality, if it could enforce net neutrality again, you know, and I think for all of the reasons that we have discussed repeatedly, including the FTC's lack of authority over common carriers, it is questionable whether they have the authority at all, but most of their authority would come from the ability to prohibit unfair and deceptive trade practices, and there is no civil penalty authority in that area.

Mr. MCNERNEY. So under Chairman Pai's plan, broadband providers are not required to disclose the practices at the point of sale or on their website, but they can give those practices to the FTC and the FCC, and they would in turn put them on their website. Is that sort of disclosure viable?

Ms. MOY. So, you know, I mean, I think I would say again, you know, I think as an initial matter it is worth remembering that the disclosures alone are not necessarily, are not going to be sufficient, particularly when it comes to when you are in a situation where a consumer only has access to one broadband provider.

But when there is a choice that is available to the consumer and they might rely on disclosures to make a choice between two different providers or between multiple providers, that information really does need to be consumer-facing. I was, in fact, on the task force at the Consumer Advisory Committee, the FCC's Consumer Advisory Committee that designed the so-called broadband nutrition label that Chairman Pai is planning to do away with, you know, and we did think that in a situation where a consumer might be considering adopting one of two different services or one of two different service plans, it would be extremely important for them to have easy-to-read information about the actual performance of that service package.

Mr. MCNERNEY. I had a couple of questions for Professor Kearns. With regard to machine learning, there are going to be benefits in all sorts of areas, but are there areas where machine-learning techniques should not be used?

Dr. KEARNS. Well, yes, I think so. And there is, you know, a large and growing community of AI and machine-learning researchers who are trying to debate those sorts of issues. You know, one logical extreme, there is the notion that any decision that really, you know, should lie with a human just because of moral agency shouldn't be made by an algorithm.

So one example that is commonly offered is in automated warfare, that even if we could design algorithms or learn models that, you know, made more accurate decisions about whether to fire on an enemy, that perhaps we shouldn't do that because the decision to do that should always lie with a human who has the moral responsibility for that decision.

So I think, you know, that is an extreme that I think I would agree with. The harder cases, I think, are cases in which, you know, machine learning is demonstrably effective yet making difficult moral decisions like in criminal sentencing and to, you know, one could arguably ask about things like, you know, college admissions or loan decisions and the like.

And so, you know, my view right now is that we are at the very beginning of a very difficult debate about the extent to which decisions that have been made historically by humans and, by the way, you know, historically also exhibited biased privacy decisions, et cetera, when they were being made by humans and turning over them to machines where the tradeoffs are going to be different, but there will be tradeoffs, right.

And there is always this tension in machine learning between accuracy, which is, you know, right now essentially what is almost always optimized for, and other things like privacy or fairness, right.

Mr. MCNERNEY. Well, I have really gone over my time.

Dr. KEARNS. OK, yes.

Mr. MCNERNEY. So I am going to have to interrupt you and yield back. Thank you.

Mr. LANCE. Thank you very much. The Chair recognizes Mr. Johnson.

Mr. JOHNSON. Thank you, Mr. Chairman.

Mr. Pasquale, when we talk about how companies interact with consumers and handle consumer data, we often talk about transparency, that is, being transparent with business practices. In some industries there are actually transparency rules that require companies to disclose certain information. For example, ISPs have to disclose a slew of information about their business and network practices. Are there any rules that require companies that use algorithms to be transparent about how they work?

Mr. PASQUALE. So it is a very narrow range of requirements. So, for example, if you look at the online lending space, there has been some caution about certain forms of automated underwriting using what is called fringe or alternative data, data beyond, you know, what is usually used by FICO or other entities like that because under FCRA it can be a requirement of explanation under the Fair Credit Reporting Act with respect to some of these, like giving the reason codes for why an automated decision was made.

But in general it is a zone of great opacity. We just don't know. That is why I titled my book "The Black Box Society," because there are so many rules there, so little sense of what is going on there. Yes.

Mr. JOHNSON. OK, all right.

Mr. Kearns, Professor Swire from Georgia Tech—my alma mater, by the way, it is where I learned about networking—concluded that applications such as search engines and social networking services collect data providing greater consumer insight than ISPs. Do you agree with that conclusion?

Dr. KEARNS. Yes, I do.

Mr. JOHNSON. OK, care to expand?

Dr. KEARNS. Well, in addition to the aforementioned encryption that, you know, occurs with the vast majority of data that ISPs

carry, you know, there is the additional fact that I don't think it has been mentioned yet that it is at the packet level. And the way internet routing, packet routing works is that longer messages, whether they are actual text messages or they are a web search or they are an audio call, are divided into these tiny little fixed-size packets which then travel possibly different paths through the network.

So, you know, just going back to a comment I made earlier, this sort of actionability of data at that level, if half or more of it is encrypted and it is also traveling in these little bite-sized pieces and you are carrying a phenomenal amount of that data over your network, if you ask me whether if I am trying to figure out who somebody is and what to sell them and what their mental and psychological condition is, I would much rather have search engine data or Facebook data than packets at the network level.

So this is basically what I mean by, I think, you know, from a privacy perspective it is less concerning to me than the data that is being collected by the edge services.

Mr. JOHNSON. OK, all right. Continuing with you, Mr. Kearns, then, my understanding is that approximately 80 percent of internet traffic is encrypted. You just talked about encryption a little bit. That limits what ISPs see regarding consumers' online activities. In contrast, by their very nature, don't edge providers largely have to interact with consumers' unencrypted data?

Dr. KEARNS. By definition, yes.

Mr. JOHNSON. Yes. Well, doesn't that give edge providers much greater insight into consumers' preferences, habits, choices, beliefs, that kind of stuff?

Dr. KEARNS. Yes, it does. I mean, I think the right way to think about it, let's say, back in the old days of telephony is, you know, would you rather see the raw analog signal and try to figure out what the conversation is from that, or would you rather have that analog signal rendered through a speaker so that you could actually listen to the conversation, right? And this is an imperfect metaphor, but I think it is a good one.

You know, another thing I might offer is, if I am just trying to describe an image to you, would you rather I go pixel by pixel through the image and tell you the color value of it, or would you rather me describe it to you and say, well, "It is an outdoor image?" "There are trees. There is a lake. There is a family picnicking." And so, you know, by definition, what the end services are getting and what users want to give to those end users are this much-higher-level data that is easy for humans to understand and model.

Mr. JOHNSON. They want to see it all put back together again.

Dr. KEARNS. Yes, exactly. And you are just kind of not easily getting that at the network level because of the encryption and because of the fragmentary nature of packet routing.

Mr. JOHNSON. Right, right. OK.

Well, Mr. Chairman, I yield back a full 10 seconds.

Mr. LANCE. Thank you, Mr. Johnson. The Chair recognizes Ms. Eshoo.

Ms. ESHOO. Thank you, Mr. Chairman, and thank you to the witnesses. I read all your testimony last night, listened to all of you today, and I want to make some comments about this hearing. The

title of it is very interesting, and it is an area that needs to be examined.

The word “privacy” has come up many times, certainly net neutrality and references to it have come up. Strong enforcement has come up. But when you look at the backdrop and the broader stage on which this hearing sits, look what is happening in our country. In a flash, like lightning, privacy was ripped away, the privacy protections were ripped away from the internet.

So all of the happy talk of some of my colleagues on this committee about privacy and the sanctity of it, that was forgotten when that vote was taken and the American consumer, I think, has really been hammered as a result of it. I think that, Professor Moy, you made a very important point when you said that the last thing we should do is to repeal, and that has happened.

It was very interesting to hear the description of what has taken place in Europe with what they have done with the internet and what we have done and how the internet has flourished in our country just on the eve of the Chairman of the Federal Communications Commission getting ready to rip away the protections that are there that have made it open, free, accessible. So I think there is some political cross-dressing here today, with all due respect, not by the panelists, but I think by some of the Members.

And the term “a strong enforcement” has been referred to, but I don’t think strong enforcement is something that you pick and choose, because we are lawmakers and, unless there is enforcement, then the law is not worth the paper that it is written on.

I take heart from what Professor Kearns spoke of because, in this whole issue of algorithms—and let’s keep in mind these social platforms are free. They are free. They are not like the ISPs. In the ISPs there must be, I think, only three happy outfits in the entire Nation on the eve of what Chairman Pai is doing relative to net neutrality, and that is Comcast, AT&T, and Verizon. They are the happiest. I don’t know anyone else that is for what he is planning to do to the internet. But I do think that it is very interesting to me that you have raised the issue of auditing algorithms.

Now, I think that truth has always required transparency. We don’t, I don’t think, as a committee, really know how to get socks on the octopus, so to speak, here because it is complicated. Free speech is central to us, but we also know that there are bad actors that have used the best of what we have invented to divide us, and something needs to be done about that. There is no question in my mind, and the chairman of the full committee raised that, as well.

So how close, Professor Kearns, do you think we are to this what you raised, the auditing of algorithms?

Dr. KEARNS. So I think we are close. So in particular, you know, many of the instances of discrimination, for instance, in algorithmic behavior were actually discovered by groups of researchers who are effectively doing their own auditing, you know, doing kind of field experiments using services that have algorithms underlying them, testing their behavior and demonstrating, for instance, they have some particular type of bias.

There is good research being done on, again, internalizing notions of fairness inside of algorithms. And just to be clear here, I think most instances of discrimination in algorithmic behavior are

not the result of any evil by the researchers and scientists at these companies. It is just that, when you optimize your model for predictive accuracy, you shouldn't expect it to have any other nice properties, either, so you need to actually specifically put those properties in your code if you want them to have it.

You know, in the privacy arena there is a very strong notion of, you know, kind of internal privacy of an algorithm known as differential privacy that is kind of starting to finally get out of the lab and be used, for instance, in the latest version of Apple's iOS. So this stuff is happening, and the tech companies are participating in, you know, the dialogue and in developing some of the science. It just needs to be kind of taken seriously at scale by those companies.

Ms. ESHOO. Well, I am encouraged by what you have just described, and I want to pursue it, as well. If there is more information that you can get to us on it, I certainly would welcome it. And with that, Mr. Chairman, I yield back.

Mr. LANCE. Thank you very much. The Chair recognizes Dr. Bucshon.

Mr. BUCSHON. Thank you, Mr. Chairman.

Professor Kearns, this is a little bit different line of questioning but important. Is it feasible for your cell phone or an app on your cell phone to listen in on your conversation and collect data?

Dr. KEARNS. Yes.

Mr. BUCSHON. And are you aware that that is happening in our country, in everywhere? I will give you an example of why I think that this is happening, and it is an issue that we really haven't touched on today as part of data collection.

Dr. KEARNS. My default assumption is that, unless I have taken explicit pains to arrange otherwise, that when I use an app on my mobile phone, it is recording at least the data of my interaction with that app and possibly many other aspects of my usage of the phone, as well.

Mr. BUCSHON. How about when you are talking? Like right now my phone is sitting here, and there is a speaker and I am talking, and is that data, is what I am saying potentially being collected?

Dr. KEARNS. With or without the microphone on.

Mr. BUCSHON. Correct, with or without. Well, the question is the definition of "on," right, because that is being made by the company that makes the phone. I mean, it has been shown recently and it has been on, I think, Wall Street or somebody reported that you can turn off essentially everything on your phone and you are still being tracked. So the speaker is important.

Let me just say this, and this is the reason this came to me is because my son, who is 24, he lives in Chicago, he was standing around with some, a couple, with a friend at work, a person at work. Nobody was on the internet. He was talking about, and I can't remember specifically what it was, but it was about shoes or something and the next day he had ads for that exact thing on his feed. He didn't do a Google, he didn't do any search. I don't want to single out a company, but he didn't do any search at all. All he did was talk in the presence of his microphone on his phone. Do we know if that is happening?

Dr. KEARNS. I am not a security expert, but I do know that there are more instances these days of situations in which, you know, the operating system on your mobile phone communicates with beacons in retail stores, and this is how one often experiences, you know, why even though I didn't do a search on some product at all, but I happened to be in the store yesterday, the physical retail store—

Mr. BUCSHON. Yes, they can do that.

Dr. KEARNS [continuing]. Am I not, you know, and this is because they are now starting to install so-called beacons in these stores that interact with the operating system on your phone, and so then the retailer knows that you were there.

Mr. BUCSHON. If you were in a shoe store, they know you were in a shoe store.

Dr. KEARNS. So, you know, my feeling about these things is that the way technology is, is anything is possible, right. And then the question is, is it widespread and who is doing it, and is it kind of for deliberately nefarious purposes or is it, you know, just advertising, quote unquote?

Mr. BUCSHON. I mean, it is important because I am a Member of Congress and I have confidential conversations all the time with my phone, and I am not on the internet. And so that is a question. I had mentioned this to my staff, by the way, when I went back to the office, and they go, "Oh yes, that has happened to me." I mean, all the young people are like, "Oh yes, that happened to me before."

So I just thought that was something that we need to, really, also as far as collecting data and then analyzing like you have described, I mean, I think what we really need to think about, not only when you are actively on your phone but whether or not through your—and I am not a conspiracy theorist or anything, right—through your actual speaker that you can be monitored.

Dr. KEARNS. Yes. I mean, and I think we are also voluntarily heading this direction in the form of home devices like, you know, Echos and, you know—

Mr. BUCSHON. Yes, right. That is obvious, right.

Dr. KEARNS [continuing]. In which, you know, are kind of sitting there all the time recording.

Mr. BUCSHON. Right.

Professor Ben-Shahar, you stated that consumers ignore privacy disclosures. How would you suggest we inform consumers that they have given consent to their data being collected? How can we do that?

Dr. BEN-SHAHAR. I think consumers understand in general what is going on, and indeed a lot of the surveys suggest that they know that a lot of their information is being collected. They are not surprised when they find out that yet another practice is prevalent, for example, that now these home butlers, the Google Home or Alexa is listening to everything that is going on. I think that consumers by now have figured out that this is going on, and so there is not much that we can tell them that they don't know.

Now there are specific things that are going on that defy consumers' expectation. And if the expectation is created in an affirmative way by your smart phone or by Google or by other service,

for example, they give the consumer the impression that they can turn on or turn off some kinds of surveillance or some kinds of data collection and it turns out that they can't, that even if they did what they were supposed to do and had the reasonable understanding that they are not going to be tracked in a particular way, they still are, that is an FTC issue. That is an issue of—

Mr. BUCSHON. Well, that has happened. It has just been written in the papers recently that it has happened.

Dr. BEN-SHAHAR. To the extent that that is happening, that should be—I think that there are tools in our law, both in contract law and in consumer protection statutory law, to take care of these kind of things. I don't know, you know, maybe other panelists know better. I don't think these things happen too much, for the simple reason that it all costs nothing for the services to let consumers know what is going on. Consumers don't care. They are not going to bother, change the settings or re-change the settings every time there is a new version of the software.

Mr. BUCSHON. Thank you. I am out of time. I yield back.

Mr. LANCE. Thank you very much.

The Chair recognizes Congressman Flores.

Mr. FLORES. Thank you, Mr. Chairman. I appreciate all the panelists for joining us for this important hearing today.

The first question I ask, I mean, one of the things that is obvious is that data is pulled from everywhere, whether it is data services, your mobile phone, your Alexa, whatever, operating systems, and social media platforms. So my question is this, for all of the panel. I am going to start with Professor Kearns, and then I am going to ask a couple of other questions, and we will come back to the panel if we have time about this issue.

So the question is simply this: What are your thoughts as to whether or not Congress or policymakers need to establish a consistent legal and regulatory framework for how this data is obtained and used?

Dr. KEARNS. Well, I will be brief so other people can talk, too. But, I mean, as per my earliest remarks, as a scientist, so I am not a policymaker, I am not a lawyer—

Mr. FLORES. Right.

Dr. KEARNS [continuing]. But from a scientific perspective, to me the most important thing is not sort of, you know, how much data you have measured in petabytes. It is not kind of whether the data came from this service or that service or this app or that ISP. It is, what are the actionable insights about consumers and what are the facts about their lives that you can infer from that data?

And as a scientist I don't see an easy way to carve that up into little subdomains and say, like, "Oh, well, you know, because we just—" the truth is, we don't know, right. These companies themselves are figuring out just now how powerful AI and machine-learning techniques applied to all kinds of data are.

Mr. FLORES. Right. Well, the challenge is, is that policymakers and regulators typically move way behind the speed of technological change. And so what I am trying to figure out is how do we get in front of this, or do we need to even worry about it? And I will come back to the rest of the panel on this question in just a

minute, but I do have two other questions for Professor Pasquale first.

In your testimony, you noted that bottlenecks can threaten competition at any layer of the network, not just the physical layer provided by the ISPs. And so the question is this: Can you elaborate on the potential bottlenecks other than the ISPs, beyond the ISPs?

Mr. PASQUALE. Sure. So I did a 2008 article called “Internet Non-discrimination Principles,” and what I tried to do is to say that the same type of concerns that are motivating people to advocate for net neutrality should also be looked at, at the social layer, at the search engine, at the app store level. And particular examples, there are two examples related to China that I think are really interesting and I discuss in my book. One is that someone developed an app called “In a Permanent Save State,” and it was a game that was also a critique of Apple and its use of certain Chinese factories and labor. And the Apple app store rejected it over and over again, and they couldn’t really understand why that was happening.

Similarly, there is a case called *Langdon v. Google* where someone wanted to buy an ad titled “China is Evil,” and there was, I thought, a relatively arbitrary decision by Google to say, “No, we are not going to sell you that ad.” And so I think those are very concrete examples of a much larger problem, where I think that we have to be much more imaginative as academics and as policymakers in seeing the connections rather than seeing the separations between these different entities.

Mr. FLORES. Well, that sort of goes to my next question, because we have talked a lot about how content is filtered online, but we need to consider how content is filtered through other platforms, even voice service devices. It has been reported that voice service devices prioritize certain content and services and they have even excluded certain products from their platforms.

So the first question is, are there anti-competitive concerns associated with this type of prioritization?

Mr. PASQUALE. Congressman Flores, I have to confess I am not familiar with that niche of the market, so I will have to pass.

Mr. FLORES. OK. That is fine. Let’s move back to my initial question, if we can. I would like to get the comments from the rest of the panel. Again, the question was this: What are your thoughts as to whether policymakers need to establish a consistent legal and regulatory framework for how this data may be obtained and used? Let’s start with Ms. Tucker.

Dr. TUCKER. So I think it is very difficult—and Europe has taught us this—to have a consistent framework governing technology. On the other hand, I think it is possible to identify areas where we are particularly concerned about privacy, be it health, be it kids, and make sure the policy is focused on protecting those outcomes we really care about.

Mr. FLORES. OK.

Dr. Ben-Shahar?

Dr. BEN-SHAHAR. My answer, with all due respect, is a resounding no. I don’t think that policymakers should tell business what data to collect and how to use it.

Mr. FLORES. In the interest of time, I appreciate the short answer, OK.

Dr. BEN-SHAHAR. And maybe just set red lines.

Mr. FLORES. Ms. Klonick, sorry.

Ms. KLONICK. Yes. I think that regulation, Section 230 and any regulation that kind of curtails the ability of these businesses and platforms to self-regulate, is probably not in the best interest of the public.

Mr. FLORES. OK, thank you.

And, in the interest of time, I will yield back the balance of my time.

Mr. LANCE. Thank you very much. The Chair recognizes Congresswoman Walters of California.

Mrs. WALTERS. Thank you. And thank you for holding this hearing, and thanks to the witnesses for being here.

We can all agree that protecting consumers' information is paramount and that consumers deserve a clear understanding of their privacy expectations when using the internet. It is important we have this discussion so we can better understand how consumers benefit from current practice and examine ways to protect against the misuse of consumer information.

Professor Tucker, what is the best way to protect my constituents' privacy to make them feel secure and confident in the use of their data without impeding future innovation and America's leadership in the technology sector?

Dr. TUCKER. So, over the various sectors and various time periods, my research has repeatedly shown that the best way of introducing privacy protections is to give a sense of control back to consumers. Now, that is distinct from transparency. It is distinct from disclosures. Instead, it is about restoring a sense of control. And what is more, my research has actually shown that that kind of policy is in from self-interest. And if you try and do the kind of microsegmentation using really personalized data, for example, preferences of someone over shoes, then using that kind of data for advertising only works if there is a parallel sense of control among consumers.

Mrs. WALTERS. OK, thank you.

Professor Ben-Shahar, what protections do existing legal schemes provide for consumers to protect them from the theft or loss of their data, and are those legal schemes sufficient?

Dr. BEN-SHAHAR. Well, I think that, again, I am not a data security expert, but my understanding is that there are very few protections that are granted to consumers. Many of the things that were recommended that people do after, for example, the Equifax breach were fairly limited. I mentioned before in my testimony that I think that the reason there are so few remedies and recourses is because largely there is no evidence for the fact that consumers are suffering in a magnitude of harm that requires greater a remedy in this context.

Mrs. WALTERS. OK. And then I have another question for you. How does the use of algorithms to deliver content impact consumers' experiences online, and is there a benefit we see to the practice of collecting data?

Dr. BEN-SHAHAR. I think that benefit is enormous, and it has been, you know, measured in many different ways. But I will just recommend to try one time to disconnect all the knowledge about

you from your smart phone and see what happens. When you open Google Maps and want to go something and it no longer recognizes after the first letter where it is that you wanted to go and the inconvenience that you will say, “Ah, no, I wish the data service was still on, the recognition was on.”

I think in many contexts personalization delivers astronomical value that has not yet been tapped. In my own research I am looking about at ways in which we can personalize legal rules and other things, but the only reason that we think about these new areas is because existing areas have proven to be enormously beneficial—education, insurance, medicine, and the like.

Mrs. WALTERS. OK, thank you.

And Professor Tucker, some digital platforms would say that, when third parties are permitted to use their platform, that platform gives consumers the tools to control their experience. Are we putting too much of the onus on the consumer to review the permissions the developer is requesting and forcing the consumer to choose which information to share?

Dr. TUCKER. So I think this is a very good distinction to make in that, let’s be clear, whenever we have actually studied search logs of how consumers behave when they are confronted by control, rather than opting out and, you know, protecting their privacy, they tend to actually go in and try and improve the data, because there is nothing more irritating—I don’t know if this has happened to you, that you are looking at a web service which thinks you are a 25-year-old man, and you are like, “Why do you think that?” Consumers tend to try and improve the quality of data, intriguingly.

The one distinction I do want to make, though, is that there are some categories of consumers where perhaps there isn’t that level of control exerted. For example, we have a study right now which looks at apps which are targeted at toddlers. I don’t know if you have ever been to a restaurant where parents are using these to quiet down their toddlers, but we saw there a vast quantity of data being collected. And there I think it is fair to assume that those toddlers are not really actually exerting any control on whether their location is being tracked or their use of the sort of My Little Pony app or whatever it is.

Mrs. WALTERS. Thank you, and I yield back the balance of my time.

Mr. LANCE. Thank you very much. The Chair recognizes Congressman Costello.

Mr. COSTELLO. Thank you. I want to share some reflections I have here and allow each of you to correct my understanding or enhance it, whatever terminology you may wish to use. From my perspective, browser history in some respects is a commodity, but it is very invisible and at this point there is no regulatory framework for when and how it can be incorporated into an algorithm.

I take, and this is not a precise corollary, but if I made a phone call to you and the content of our discussion was transcribed and it was then sold or utilized for proprietary or commercial gain, there are some similarities between that and how an ISP is able to gather some of that content and then incorporate that into an algorithm or into how advertising would make its way into my

internet searches, or if I go to a news website, all of a sudden up pops laundry detergent if I was Googling laundry detergent.

Someone made the comment about editorializing content or raise concerns on the political side. It may have been Ms. Moy in her written testimony. I read everyone's written testimony. The trouble, the thing that I am grappling with on the concerns related to what kind of political content shows up and how you might be able to shape one's opinion of things is, what is the difference between that and picking up a newspaper in the morning? And I don't really know how to distinguish between—you can distinguish between the two, but in some respects I don't know that you should distinguish between the two.

As it relates to the Federal Trade Commission, if we are talking about, particularly on political content, but even amongst other things, how would the FTC go about adjudicating equal time if we were to get into talking about political content, and how does it get, how do you determine, oh, well, you put too much left-leaning or too much right-leaning content? I think that that can get deeply problematic.

And I believe, also, Ms. Moy mentioned something about adding protections for consumers, if you could share with me what kind of protections you might be speaking about.

The gentleman, I believe it was Dr. Ben-Shahar, I agree with your testimony. I don't think that these waivers or disclaimers or—it doesn't mean a hill of beans. I totally agree with you. I am not sure, I think that is just more about indemnification or protecting one's liability, and that is fine. I mean, I don't think we should expect more from that. I don't know how you could expect more from it.

But the final thing I want to say for comment relates to Ms. Tucker's testimony. And in the final two paragraphs, you talk about how different types of data can have different consequences and that any regulation, rather than treating all the data the same, needs to distinguish between what kinds of data may be actively harmful to consumers and what data may not be.

And it seems to me that we are really talking about values here, right. We want algorithms to be able to be helpful to the consumer and, candidly, in some respects helpful to those who are going to use that data to make sure that you have information that you may be more predisposed to wanting to see. We don't want that data to be harmful.

See, I am going on way too long. How do we create a clear yet evaluative standard and entrust everyone to follow it with enough tools for the FTC to embrace that kind of framework if we were to do it? I have spoken way too long. Comments?

Mr. PASQUALE. I mean, I just want to—I have two quick responses, one being that I do think that, you know, in terms of thinking about what data is sensitive and what is not, that can be a strength of a privacy regime.

But if we also look at the work on big data proxies, how like Nicholas Terry has described, how you can have, say, location could be a proxy for race or the very data that you don't think is terribly sensitive could be a proxy for other data that is sensitive, that is where I would turn sort of Dr. Kearns' work against Dr. Tucker's

work in a way and sort of say that there is a way in which, you know, it is because of these sort of inferences you can make from somewhat insensitive data to sensitive that is important.

With respect to Google and the newspaper, the difference that I would make is that I would say that what we are concerned often with respect to unfair algorithmic influence on political activity would be something that was a lot more subtle. So, for example, imagine if Facebook decided it was only going to encourage Democrats to vote. We do have studies that have shown that that can lead to I think it is a 0.63 or a small increase on the margin of the people whose feed is spiced with get-out-the-vote advertisements.

So that is something I think we definitely have to look for because, when a newspaper says “Vote for X,” I can see that. But when Facebook, you know, suddenly spices the feed of the people that, say, it likes, then we can’t see that.

Mr. COSTELLO. Fair point.

Ms. MOY. So yes, and I will just add, you know, when it comes to—so a couple things. One, you know, when it comes to the FTC’s enforcement authority, at the risk of sounding like a broken record, the enforcement authority really is limited to deception, unfair and deceptive practices, and there is no civil penalty authority.

But, you know, on your question of paid political ads, specifically, you know, I think that this is a really hard challenge that I suspect we don’t have a lot of really good answers for yet on how to deal with. You know, one thing, though, is that there is very little transparency about what ads are being paid for and even when they contain political content. The FEC is conducting a rulemaking right now to at least explore the possibility of increasing transparency when it comes to labeling of political content on platforms, but—
or online, I should say.

You know, but I think also this is a question where it might be extremely difficult to identify some political content, for example, when it relates to issues as opposed to candidates, without human eyeballs. And there is a tremendous amount of content that gets posted online and not nearly enough human eyeballs reviewing some of that content to determine whether and to what extent it might have a political effect.

Mr. COSTELLO. I am just going to read this, something real quick into the record. I know you are ready to get out of here, Mr. Chairman. When someone states, quote, “I could slow down”—well, we talked a lot about power that exists in the hands of those that are not ISPs. For instance, just last weekend, Matthew Prince, the CEO of Cloudflare, signaled he would look into taking up a challenge to slow down the FCC Chairman’s internet speed at his home. These apparently are not the least of the threats to Chairman Pai’s home life.

When someone states, quote, “I could do this in a different but equally effective way”—and I would like to submit the entire string of tweets for the record—isn’t it clear there is a great deal of power in those that are not governed by the same rules in the internet ecosystem? And how would your reaction be different if an ISP did this rather than an edge provider?

We don't have time, but if we could take any comments for the record on that, because we are dealing with this larger net neutrality issue, and I think some of the concerns are that it is not just ISPs that we should be looking at. There are some others that aren't governed that clearly have the power to do things that we all have concerns about. I yield back.

[The information appears at the conclusion of the hearing.]

Mr. LANCE. Thank you very much, Congressman Costello.

Seeing there are no further Members wishing to ask questions, I thank all of our witnesses for being here today. Before we conclude, I include the following documents to be submitted for the record by unanimous consent: a paper from the 21st Century Privacy Coalition, a letter from the Electronic Privacy Information Center.¹

Pursuant to committee rules, I remind Members that they have 10 business days to submit additional questions for the record and I ask that witnesses submit their response within 10 business days upon receipt of the questions. Without objection, the subcommittee is adjourned.

[Whereupon, at 12:47 p.m., the subcommittee was adjourned.]

[Material submitted for inclusion in the record follows:]

¹The paper has been retained in committee files and also is available at <http://docs.house.gov/meetings/IF/IF17/20171129/106659/HHRG-115-IF17-20171129-SD004-U4.pdf>. The letter appears at the conclusion of the hearing.

Congress of the United States
Washington, DC 20515

November 1, 2016

Dear Mr. Zuckerberg,


We are writing to express our deep concerns with reports that Facebook's "Ethnic Affinities" advertising customization feature allows for advertisers to exclude specific racial and ethnic groups when placing housing advertisements. This is in direct violation of the Fair Housing Act of 1968, and it is our strong desire to see Facebook address this issue immediately.

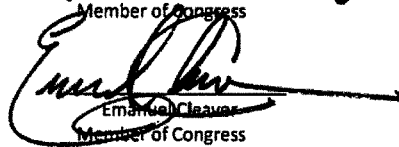
Under the Fair Housing Act of 1968, it is illegal "to make, print, or publish, or cause to be made, printed, or published any notice, statement, or advertisement, with respect to the sale or rental of a dwelling that indicates any preference, limitation, or discrimination based on race, color, religion, sex, handicap, familial status, or national origin" (42 U.S.C. § 3604). By allowing online advertisers to promote or market a community or home for the purpose of sale to select an "ethnic affinity" as part of their advertising campaign, Facebook is complicit in promoting restrictive housing practices.


It is our sincere hope that the advent of this customization microtargeting feature was to be innovative and efficient, and that Facebook did not wittingly create this feature with the purpose of separating communities or violating federal civil rights law. That said, in light of this revelation, it is your responsibility as Facebook's Chief Executive Officer to remedy this matter swiftly and responsibly. On a similar note, with 2 percent of Facebook's U.S. employees being African American, and 4 percent Hispanic, we remain convinced that a stronger commitment to diversifying the ranks of your company, especially in senior management positions to better reflect the diversity of your 1.7 billion monthly users will help in ensuring that innovative and inclusive platforms continue to be promoted by your company. Additionally, programs or policies that are potentially violative of civil rights laws or racially insensitive have an even greater likelihood of being preempted as they will be subject to a more robust and inclusive vetting process.

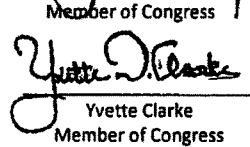
We ask that you provide a timely response to this letter outlining the scope of use of the "Ethnic Affinity" feature in housing advertisement on your site, and what steps – if any – are being made to ensure that Facebook is not empowering discriminatory housing practices. Please do not hesitate to contact us in the interim as you address this matter. We look forward to a constructive dialogue on Facebook's future innovation, inclusive advertising, and efforts to diversify the tech sector.

Regards,


Robin L. Kelly
Member of Congress


Emanuel Cleaver
Member of Congress


G.H. Butterfield
Member of Congress


Yvette Clarke
Member of Congress

Written Testimony of
Frank Pasquale
Professor of Law
University of Maryland

Before the United States Senate
Committee on the Banking, Housing, and Urban Affairs

“Exploring the Fintech Landscape”
Sept. 12, 2017
10:00 am
Dirksen Senate Office Building

Witness Background

Frank Pasquale is Professor of Law at the University of Maryland's Francis King Carey School of Law. His research addresses the challenges posed to law by rapidly changing technology. He has served as a member of the NSF-funded Council for Big Data, Ethics, and Society, and is an Affiliate Fellow of Yale Law School's Information Society Project. His recent publications focus on the legal implications of big data, artificial intelligence, and algorithms. His book *The Black Box Society: The Secret Algorithms that Control Money and Information* (Harvard University Press, 2015) has informed the global algorithmic accountability movement, and has been translated into Chinese, French, and Serbian. He is a co-founder of the Association for the Promotion of Political Economy and Law (APPEAL), with Jennifer Taub and Martha McCluskey.

Pasquale has been a Visiting Fellow at Princeton's Center for Information Technology Policy, and a Visiting Professor at Yale Law School and Cardozo Law School. He serves on the Advisory Boards of the Data Competition Institute, the New Economy Law Center, and the Electronic Privacy Information Center. He has co-authored a casebook on administrative law and co-authored or authored over 50 scholarly articles, including *Law's Acceleration of Finance: Redefining the Problem of High-Frequency Trading*, 36 *Cardozo Law Review* 2085 (2015); *Four Futures of Legal Automation*, 63 *UCLA Law Review Discourse* 26 (2015) (with Glyn Cashwell); *The Scored Society: Due Process for Automated Predictions*, 89 *Washington Law Review* 1 (2014) (with Danielle Citron); *The Troubling Consequences of Trade Secret Protection of Search Engine Rankings*, in *The Law and Theory of Trade Secrecy* (Rochelle Cooper Dreyfuss & Katherine Jo Strandburg eds., 2011); and *Democratizing Higher Education: Defending & Extending Income Based Repayment Programs*, 28 *Loyola Consumer Law Review* 1 (2015). He graduated with a B.A., *summa cum laude*, from Harvard University, an MPhil. from Oxford, and a JD from Yale Law School.

I. Introduction

The financial technology (“fintech”) landscape is complex and diverse. Fintech ranges from automation of office procedures once performed by workers, to some genuinely new approaches to storing and transferring value, and granting credit.¹ New services—like insurance sold by the hour—are emerging. Established and start-up firms are using emerging data sources and algorithms to assess credit risk. And even as financial institutions are adopting some distributed ledger technologies, some proponents of cryptocurrency claim that it “changes everything” and will lead to a “blockchain revolution.”

For purposes of this testimony, I will divide the fintech landscape into two spheres. One, incrementalist fintech, uses new data, algorithms, and software to perform classic work of existing financial institutions. This new technology does not change the underlying nature of underwriting, payment processing, lending, or other functions of the financial sector. Regulators should, accordingly, assure that long-standing principles of financial regulation persist here. I address these issues in Part II below.

Another sector, which I deem “futurist fintech,” claims to disrupt financial markets in ways that supersede regulation, or render it obsolete. For example, if you truly believe a blockchain memorializing transactions is “immutable,” you may not see the need for regulatory interventions to promote security to stop malicious hacking or modification of records. In my view, futurist fintech faces fundamental barriers to widespread realization and dissemination. I address these issues in Part III below.

II. Incrementalist Fintech

A. Big Data or Artificial Intelligence-based Underwriting

Many marketplace lenders are now using forms of data not traditionally used for credit underwriting, in order to offer consumer or small business loans. They may help correct some long-standing problems in US credit markets, including the problematic nature of contemporary credit scoring. However, as Mikella Hurley & Julius Adebayo have argued,

Credit-scoring tools that integrate thousands of data points, most of which are collected without consumer knowledge, create serious problems of transparency.

¹ The Government Accountability Office has described fintech as follows: “The financial technology (fintech) industry is generally described in terms of subsectors that have or are likely to have the greatest impact on financial services, such as credit and payments. Commonly referenced subsectors associated with fintech include marketplace lending, mobile payments, digital wealth management, and distributed ledger technology.” GAO, FINANCIAL TECHNOLOGY: INFORMATION ON SUBSECTORS AND REGULATORY OVERSIGHT (2017).

Consumers have limited ability to identify and contest unfair credit decisions, and little chance to understand what steps they should take to improve their credit. Recent studies have also questioned the accuracy of the data used by these tools, in some cases identifying serious flaws that have a substantial bearing on lending decisions.

Big-data tools may also risk creating a system of "creditworthiness by association" in which consumers' familial, religious, social, and other affiliations determine their eligibility for an affordable loan. These tools may furthermore obscure discriminatory and subjective lending policies behind a single "objective" score. Such discriminatory scoring may not be intentional; instead, sophisticated algorithms may combine facially neutral data points and treat them as proxies for immutable characteristics such as race or gender, thereby circumventing existing non-discrimination laws and systematically denying credit access to certain groups. Finally, big-data tools may allow online payday lenders to target the most vulnerable consumers and lure them into debt traps.²

The problem of "big data proxies" is a serious one recognized by leading privacy scholars.³ Regulators should do much more to assure that next-generation technology does not simply reproduce old biases.⁴ The alternative is a "scored society" where individuals lack basic information about how they have been treated in the credit granting context.⁵

These problems are troubling in the abstract. Their concrete implications are chilling, as a recent Privacy International Report revealed. Outside the United States, fintech firms have already scored creditworthiness based on the following factors:

- "If lenders see political activity on someone's Twitter account in India, they'll consider repayment more difficult and not lend to that individual."
- "The contents of a person's smartphone, including who and when you call and receive messages, what apps are on the device, location data, and more."
- "How you use a website and your location. [One firm] analyses the way you fill in a form (in addition to what you say in the form), and how you use a website, on what kind of device, and in what location."⁶

² Mikella Hurley & Julius Adebayo, *Credit Scoring the Era of Big Data*, 18 YALE J.L. & TECH. 148 (2017).

³ See, e.g., Nicolas Terry, *Big Data Proxies and Health Privacy Exceptionalism*, HEALTH MATRIX (2015).

⁴ For an up-to-the-minute overview of this and related problems, see Penny Crosman, *Is AI a threat to fair lending?*, at <https://www.americanbanker.com/news/is-artificial-intelligence-a-threat-to-fair-lending>.

⁵ Danielle Keats Citron & Frank Pasquale, *The Scored Society*, 89 WASH. L. REV. 1 (2014).

⁶ Privacy International, *Case Study: Fintech and the Financial Exploitation of Customer Data*, at

Moreover, machine learning systems are constantly developing even more invasive forms of assessing creditworthiness, or factors influencing it. A recently published paper claims to infer propensity to criminality merely from the features of persons' faces.⁷ Sexuality and health are also now being predicted by machine learning researchers entirely on the basis of a picture of a person's face—something relatively easy to gather via a Google image search, or Facebook search.⁸ Regulators need to be able to audit machine learning processes to understand, at a minimum, whether suspect sources of data like these are influencing fintech firms.⁹

1. Neither Machine Learning Nor Predictive Analytics are too Complex to Regulate

Some fintech firms which rely on artificial intelligence may counter that the computation involved in their decisionmaking now amounts to a form of cognition as hard to explain as that of a human decision-maker. Genetic algorithms may, for instance, themselves spawn, each second, dozens of ways of processing information, which are then evaluated on some metric, and Darwinianly given a chance to persist based on their performance. Iterative machine learning processes may be similarly complex and opaque. Their view is that, just as we can't map all the brain's neurons to connect a person's decision to eat a slice of cake to some set of synapses, we can't map or unravel the sequence of events that leads to a given algorithmic score or sorting.

I believe that we should be suspicious of the deregulatory impulse behind characterizations of machine learning as “infinitely complex,” beyond the scope of human understanding. The artificial intelligence that commercial entities celebrate can just as easily evince artificial imbecility, or worse. Moreover, there are several practical steps we can take even if machine learning processes are extraordinarily complex.

<https://privacyinternational.org/node/1499?PageSpeed=noscript> (Aug. 30, 2017). See also Josh Chin & Gillian Wong, *China's New Tool for Social Control: A Credit Rating for Everything*, Wall St. J., Nov. 28, 2016, at <https://www.wsj.com/articles/chinas-new-tool-for-social-control-a-credit-rating-for-everything-1480351590>; Ian Bogost, *Cryptocurrency Might be a Path to Authoritarianism*, The Atlantic, May 30, 2017, at <https://www.theatlantic.com/technology/archive/2017/05/blockchain-of-command/528543/>.

⁷ Blaise Agüera y Arcas, Margaret Mitchell and Alexander Todorov, *Physiognomy's New Clothes*, at <https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a> (May 6, 2017).

⁸ Sam Levin, LGBT groups denounce 'dangerous' AI that uses your face to guess sexuality, *The Guardian*, at <https://www.theguardian.com/world/2017/sep/08/ai-gay-gaydar-algorithm-facial-recognition-criticism-stanford>, Sept. 8, 2017; Barbara Marquand, *How Your Selfie Can Affect Health Insurance*, USA Today, at <https://www.usatoday.com/story/money/personalfinance/2017/04/25/how-your-selfie-could-affect-your-life-insurance/100716704/>.

⁹ To be clear, I am not alleging any particular fintech firm in the United States is using such approaches in the United States at present. I am just pointing out that the possibility exists, and must be monitored.

For example, we may still want to know what data was fed into the computational process. Presume as complex a credit scoring system as possible. Regulators could still demand to know the data sets fed into it, and, for example, forbid health data from being included in that set. We already know that at least one credit card company has paid attention to certain mental health events, like going to marriage counseling.¹⁰ When statistics imply that couples in counseling are more likely to divorce than couples who aren't, counseling becomes a "signal" that marital discord may be about to spill over into financial distress.¹¹ This is effectively a "marriage counseling penalty," and poses a dilemma for policy makers. Left unrevealed, it leaves cardholders in the dark about an important aspect of creditworthiness. Once disclosed, it could discourage a couple from seeking the counseling they need to save their relationship.

There doesn't have to be any established causal relationship between counseling and late payments; correlation is enough to drive action. That can be creepy in the case of objectively verifiable conditions, like pregnancy. And it can be devastating for those categorized as "lazy," "unreliable," "struggling," or worse. Runaway data can lead to *cascading disadvantages* as digital alchemy creates new analog realities.¹² Once one piece of software has inferred that a person is a bad credit risk, a shirking worker, or a marginal consumer, that attribute may appear with decision-making clout in other systems all over the economy. There is also little in current law to prevent companies from selling their profiles of consumers.¹³

2. The Problems of Extant Data Collectors are a Reason for More Scrutiny of Fintech, Not Less

Having eroded privacy for decades, shady, poorly regulated data miners, brokers and resellers have now taken creepy classification to a whole new level. They have created lists of victims of sexual assault, and lists of people with sexually transmitted diseases. Lists of people who have Alzheimer's, dementia and AIDS. Lists of the impotent and the depressed.

¹⁰ Charles Duhigg, "What Does Your Credit Card Company Know about You?" *New York Times*, May 17, 2009, <http://www.nytimes.com/2009/05/17/magazine/17credit-t.html?pagewanted=all>. For a compelling account for the crucial role that the FTC plays in regulating unfair consumer practices and establishing a common law of privacy, see Daniel J. Solove and Woodrow Hartzog, "The FTC and the New Common Law of Privacy," *Columbia Law Review* 114 (2014): 583–676.

¹¹ Charles Duhigg, "What Does Your Credit Card Company Know about You?", *N.Y. Times*, May 12, 2009.

¹² Cathy O'Neil, *Weapons of Math Destruction* (2016).

¹³ Kashmir Hill, "Could Target Sell Its 'Pregnancy Prediction Score'?" *Forbes*, February 16, 2012, <http://www.forbes.com/sites/kashmirhill/2012/02/16/could-target-sell-its-pregnancy-prediction-score/>.

There are lists of “impulse buyers.” Lists of suckers: gullible consumers who have shown that they are susceptible to “vulnerability-based marketing.” And lists of those deemed commercially undesirable because they live in or near trailer parks or nursing homes. Not to mention lists of people who have been accused of wrongdoing, even if they were not charged or convicted. Typically sold at a few cents per name, the lists don’t have to be particularly reliable to attract eager buyers. And there is increasing risk that your spouse, friends, boss, or acquaintances could buy such data.¹⁴

There are three problems with these lists. First, they are often inaccurate. For example, as The Washington Post reported, an Arkansas woman found her credit history and job prospects wrecked after she was mistakenly listed as a methamphetamine dealer. It took her years to clear her name and find a job.¹⁵ Second, even when the information is accurate, many of the lists have no business being in the hands of fintechs. Having a medical condition, or having been a victim of a crime, should not be part of credit decisions, since such data use generates risk of compounding, self-reinforcing disadvantage via digital stigma.

Third, people aren’t told they are on these lists, so they have no opportunity to correct bad information. The Arkansas woman found out about the inaccurate report only when she was denied a job. She was one of the rare ones. The market in personal information offers little incentive for accuracy; it matters little to list-buyers whether every entry is accurate — they need only a certain threshold percentage of “hits” to improve their targeting. But to individuals wrongly included on derogatory lists, the harm to their reputation is great.¹⁶

The World Privacy Forum, a research and advocacy organization, estimates that there are about 4,000 data brokers. They range from publicly traded companies to boutiques. Companies like these vacuum up data from just about any source imaginable: consumer health websites, payday lenders, online surveys, warranty registrations, Internet sweepstakes, loyalty-card data from retailers, charities’ donor lists, magazine subscription lists, and information from public records.

It’s unrealistic to expect individuals to inquire, broker by broker, about their files. Instead, we need to require brokers to make targeted disclosures to consumers. Uncovering

¹⁴ Theodore Rostow, *What Happens When an Acquaintance Buys Your Data?: A New Privacy Harm in the Age of Data Brokers*, 34 YALE JOURNAL ON REGULATION (2016).

¹⁵ Yan Q. Mi, *Little-known firms tracking data used in credit scores*, WASH. POST, July 16, 2011, at https://www.washingtonpost.com/business/economy/little-known-firms-tracking-data-used-in-credit-scores/2011/05/24/gjQAXHcWII_story.html?utm_term=.db2a64c53efd.

¹⁶ Note that information generated for or within a credit context may spread outside it—and vice versa. Amy Traub, *Discredited: How Employment Credit Checks Keep Qualified Workers Out of a Job* (2012), <http://www.demos.org/discredited-how-employment-credit-checks-keep-qualified-workers-out-job>. Such data and inferences are very important

problems in Big Data (or decision models based on that data) should not be a burden we expect individuals to solve on their own.

Privacy protections in other areas of the law can and should be extended to cover the consumer data now fueling fintech underwriting. The Health Insurance Portability and Accountability Act, or HIPAA, obliges doctors and hospitals to give patients access to their records. The Fair Credit Reporting Act gives loan and job applicants, among others, a right to access, correct and annotate files maintained by credit reporting agencies.

It is time to modernize these laws by applying them to all companies that peddle sensitive personal information. If the laws cover only a narrow range of entities, they may as well be dead letters. For example, protections in HIPAA don't govern the "health profiles" that are compiled and traded by data brokers or fintech firms, which can learn a great deal about our health even without access to medical records.

Congress should require data brokers to register with the Federal Trade Commission, and allow individuals to request immediate notification once they have been placed on lists that contain sensitive data. Reputable data brokers will want to respond to good-faith complaints, to make their lists more accurate. Plaintiffs' lawyers could use defamation law to hold recalcitrant firms accountable.

We need regulation to help consumers recognize the perils of the new information landscape without being overwhelmed with data. The right to be notified about the use of one's data and the right to challenge and correct errors is fundamental. Without these protections, we'll continue to be judged by a big-data Star Chamber of unaccountable decision makers using questionable sources.

Policymakers are also free to restrict the scope of computational reasoning too complex to be understood in a conventional narrative or equations intelligible to humans. They may decide: if a bank can't give customers a narrative account of how it made a decision on their loan application, including the data consulted and algorithms used, then the bank can't be eligible for (some of) the array of governmental perquisites or licenses so common in the financial field. They may even demand the use of public credit scoring models, or fund public options for credit. Finally, they should look to Europe's General Data Protection Regulation (GDPR), which provides several standards for algorithmic accountability.¹⁷

¹⁷ See, e.g., Bryce W. Goodman, *A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection*, at <http://www.mlandthelaw.org/papers/goodman1.pdf> ("If implemented properly, the algorithm audits supported by the GDPR could play a critical role in making algorithms less discriminatory and more accountable.").

B. Emerging Issues in Preemption and Regulatory Arbitrage

Some fintech advocates advocate radical deregulation of their services, to enable their rapid entry into traditional banking markets. However, there is a risk of the fintech label merely masking “old wine in new bottles.” The annals of financial innovation are long, but not entirely hallowed.¹⁸ When deregulatory measures accelerated in the late 1990s and early 2000s, their advocates argued that new technology would expertly spread and diversify risk. However, new quantitative approaches often failed to perform as billed. Most fundamentally, a technology is only one part of a broader ecosystem of financial intermediation.¹⁹

I do believe that some fintech may promote competition and create new options for consumers. But we should ensure that it is fair competition, and that these options don’t have hidden pitfalls. In my research on the finance and internet sectors, I have explored patterns of regulatory arbitrage and opaque business practices that sparked the mortgage crisis of 2008.²⁰ I see similar themes emerging today.

In the run-up to the crisis, federal authorities preempted state law meant to protect consumers.²¹ The stated aim was to ensure financial inclusion and innovation, but the unintended consequences were disastrous. Federal authorities were not adequately staffed to monitor, let alone deter or punish, widespread fraudulent practices. Agencies like the Office of the Comptroller of Currency (OCC) also flattened diverse state policies into a one-size-fits-all, cookie-cutter approach. We all know the results.²² It now appears that the OCC may be repeating its past mistakes.

¹⁸ FINANCIAL CRISIS INQUIRY COMMISSION, FINAL REPORT OF THE NATIONAL COMMISSION ON THE CAUSES OF THE FINANCIAL AND ECONOMIC CRISIS IN THE UNITED STATES (2011)

¹⁹ Tom C.W. Lin, *Infinite Financial Intermediation*, 50 WAKE FOREST L. REV. 643 (2015). This article’s sections on “linked stability,” “financial cybersecurity,” and “intermediary independence” (pages 661 onwards) should be of particular interest to the committee. See also Tom C.W. Lin, *The New Financial Industry*, 65 ALA. L. REV. 567, 595 ff. (2014) (offering 10 “regulatory principles for the new financial industry”).

²⁰ Frank Pasquale, *The Black Box Society* (2015). Chapter 4 (*Finance’s Algorithms: The Emperor’s New Codes*) describes these problems in detail. Chapter 5 offers regulatory proposals.

²¹ FCIC Report, 112 and *passim* (“Once OCC and OTS preemption was in place, the two federal agencies were the only regulators with the power to prohibit abusive lending practices by national banks and thrifts and their direct subsidiaries.”); *id.*, at 350 (“The Office of Thrift Supervision has acknowledged failures in its oversight of AIG. . . a former OTS director[] told the FCIC that as late as September 2008, he had “no clue—no idea—what [AIG’s] CDS liability was.”).)

²² Fortunately, the Supreme Court quickly signalled after the crisis that its pro-preemption approach here had gone too far. See Arthur E. Wilmarth, *Cuomo v. Clearing House: The Supreme Court Responds to the Subprime Financial Crisis and Delivers a Major Victory for the Dual Banking System and Consumer Protection*, in THE PANIC OF 2008: CAUSES, CONSEQUENCES AND IMPLICATIONS FOR REFORM, Lawrence E. Mitchell and Arthur E. Wilmarth Jr., eds., Edward Elgar Publishing, 2010, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1499216.

The OCC has released a White Paper, Exploring Special Purpose National Bank Charters for Fintech Companies, in 2016 (“White Paper”).²³ The OCC believes that such charters “could advance important policy objectives, such as enhancing the ways in which financial services are provided in the 21st century, while ensuring that new fintech banks operate in a safe and sound manner, support their communities, promote financial inclusion, and protect customers.”²⁴ The OCC is, to be sure, well-intentioned. Its Office of Innovation has energetically helped entrepreneurs to understand regulatory mandates by offering informal, candid discussions “with OCC staff regarding financial technology, new products or services, partnering with a bank or fintech, or any other matter related to financial innovation.”²⁵ However, several negative consequences could arise out of OCC efforts to go beyond informal counseling about extant legal obligations, by substantively altering these obligations via special purpose national bank charters for fintech firms.

For example, such fintech charters could enable regulatory arbitrage around state restrictions on payday lending. As 270 entities--community, labor, civil rights, faith-based, and military and veterans groups--observed earlier this year, 90 million Americans “live in jurisdictions where payday lending is illegal.”²⁶ These state consumer protection laws help consumers “save billions of dollars each year in predatory payday loan fees that trap people in long-term, devastating cycles of debt.”²⁷ OCC should not take action to preempt them.²⁸

²³ Office of the Comptroller of the Currency, Exploring Special Purpose National Bank Charters for Fintech Companies (2016), <https://www.occ.treas.gov/topics/bank-operations/innovation/special-purposenational-bank-charters-for-fintech.pdf> (“White Paper”).

²⁴ *Id.*, at 2.

²⁵ OCC Office of Innovation Office Hours, at, e.g., <https://www.occ.gov/topics/responsible-innovation/innovation-office-hours.pdf>; see also CFPB’s Project Catalyst.

²⁶ Center for Responsible Lending, States without Payday and Car-title Lending Save Over \$5 Billion in Fees Annually, at

http://www.responsiblelending.org/sites/default/files/nodes/files/research-publication/crl_payday_fee_savings_jun2016.pdf (2016); Comment Letter of Over 200 Community, Labor, and Nonprofit Groups, at http://www.neweconomynyc.org/wp-content/uploads/2017/01/comment_occ_fintech_01132017.pdf (2017) (“While the fintech industry has the potential to encourage innovation, we have also seen costly payday lenders hide behind the costume of ‘fintech.’”).

²⁷ *Id.*

²⁸ Americans for Financial Reform, Exploring Special Purpose National Bank Charters for Fintech Companies, Comment Letter, Jan. 15, 2017, at <https://www.occ.gov/topics/responsible-innovation/comments/comment-americans-for-financial-reform.pdf> (explaining broad array of legal and policy concerns that would arise if such charters were granted); Center for Digital Democracy and U.S. PIRG, Exploring Special Purpose National Bank Charters for Fintech Companies, Comment Letter, at <https://www.occ.gov/topics/responsible-innovation/comments/comment-cdd-uspirg.pdf> (“lack of transparency around the processing of data and automated algorithms may lead to increasing information asymmetries between the financial institution and the individual and thus consumers are left with less awareness and a lack of understanding and control over important financial decisions.”).

These are not mere hypothetical concerns; as the New Economy Project has documented, online lenders “have been subject to a long list of state and federal enforcement actions, settlement agreements, and investigations.”²⁹ Moreover, they may lure unsuspecting borrowers away from much more sustainable alternatives, including publicly vetted options.³⁰

Nor should the Senate rush to consider a proposed bill to legislatively overturn the 2nd Circuit’s decision in *Madden v. Midland Funding, LLC*, which applied New York state usury law to loans purchased by a debt collector who believed that those laws would be preempted, since the loans were originated by a national bank.³¹ As Adam Levitin has explained, there are not sound legal or policy arguments to ground present challenges to *Madden*.³² As Levitin explains, “Preemption is part of a package with regulation, but once the loan passes beyond the hands of a National Bank, it loses its preemption protection and becomes subject to state usury laws.”³³ There is little reason to undermine the dual banking system by applying a talismanic shield against usury laws to loans even once they have been sold by the intended beneficiary of preemption.³⁴

One more aspect of regulatory arbitrage is now in fintech news: recent applications by Square and SoFi for Industrial Loan Company (ILC) charters. Walmart’s 2006

²⁹ New Economy Project, Testimony Of New Economy Project Before The New York Senate Committees On Banks And Consumer Protection and the Assembly Committees On Banks, Small Business, and Consumer Affairs & Protection, Public Hearing on Online Lending Practices, at <http://www.neweconomynyc.org/resource/testimony-nys-senate-assembly-hearing-regarding-online-lending/>. For more on New York concerns, see Daniel Alter, The “Business of Banking” in New York – An Historical Impediment To the OCC’s Proposed National “Fintech Charter,” Notice & Comment, Blog of the Yale J. Reg., June 29, 2017, at <http://yalejreg.com/nc/the-business-of-banking-in-new-york-an-historical-impediment-to-the-occs-proposed-national-fintech-charter-by-daniel-s-alter/>.

³⁰ David Lazarus, Pricey ‘fintech’ lenders put the squeeze on cash-strapped small businesses, LA Times, June 16, 2017, at <http://www.latimes.com/business/lazarus/la-fi-lazarus-small-business-loans-20170616-story.html> (reporting that an “associate administrator for the federal Small Business Administration’s Office of Capital Access, advised starting the hunt for capital not with a fintech firm but with the agency’s LINC search tool (that’s LINC as in Leveraging Information and Networks to access Capital),” in response to Lazarus’s story of a small business owner charged amounts that “translated to an annual percentage rate of 55%” by a fintech firm).

³¹ *Madden v. Marine Midland Funding*, No. 14-2131 (2d Cir. 2015).

³² Adam Levitin, *Madden v. Marine Midland Funding*, <http://www.creditslips.org/creditslips/2015/07/madden-v-marine-midland-funding.html>.

³³ *Id.*; see also Adam Levitin, *Hydraulic Regulation: Regulating Credit Markets Upstream*, 26 Yale Journal on Regulation (2009).

³⁴ Adam Levitin, *Guess Who’s Supporting Predatory Lending*, Credit Slips, <http://www.creditslips.org/creditslips/2017/08/guess-whos-supporting-predatory-lending.html> (2017) (“[T]here’s no problem with the world post-Madden, so why mess with things. But if a “fix” is needed, it ought to be (1) narrowly tailored, and (2) ensure maximum consumer protection. . . . [A]ny fix that goes beyond protecting securitizations by banks in which servicing is retained is facilitating predatory lending.”).

application for an ILC charter was eventually withdrawn, but it led to a compelling policy argument about the optimal separation between banking and commerce.³⁵ Arthur E. Wilmarth, Jr., warned that allowing commercial firms to acquire ILCs would conflict with the general American financial policy of separating banking and commerce, generate systemic risk, and enable the resulting ILCs and their parent firms to avoid necessary regulatory scrutiny, since “FDIC does not have authority to exercise consolidated supervision over commercial owners of ILCs.”³⁶ Professor Mehra Baradaran countered that, in some instances, allowing firms to merge banking and commerce functions could enhance the safety and soundness of the banking system.³⁷

However, in this case, neither SoFi nor Square appear to be the type of commercial firms which would fit Baradaran’s account, since they would not inject the source of strength that was praised by Baradaran in the Walmart scenario (a large and viable non-financial business) into the banking system. I agree with Professor Wilmarth that “Banking-industrial combinations would . . . create unfair competitive advantages for large commercial and industrial firms that can afford the costs of acquiring and operating banks.”³⁸ Far more study of fintech as a sector is needed before the FDIC grants such applications. As Rep. Maxine Waters has observed, in a detailed letter to the FDIC calling for a public hearing on the issue, premature granting of applications for ILCs “would set a precedent that a wide variety of other fintech companies may choose to follow even though concerns related to financial inclusion, consumer benefits, supervision, and regulation of such entities are still unresolved.”³⁹

The Fed was right to call for the closure of the ILC loophole last year. Though there was an interesting scholarly debate after WalMart applied to obtain an ILC charter in 2006, some more recent, post-moratorium applicants do not appear to have the redeeming characteristics of a large commercial firm. They could also be acquired by other firms, further eroding the division between banking and commerce that lies at the heart of U.S. financial regulatory goals. As Professor Wilmarth has argued, given high concentration levels in the economy in general, and the technology sector in particular, “If we permit the formation of new banking-industrial conglomerates, we will be putting more of our eggs

³⁵ WalMart and several other commercial firms applied to acquire ILCs from 2005-2006.

³⁶ Arthur E. Wilmarth, Jr., *Wal-Mart and the Separation of Banking and Commerce*, 39 Conn. L. Rev. 1539 (2007).

³⁷ Mehra Baradaran, *Reconsidering the Separation of Banking and Commerce*, 80 George Washington Law Review 385 (2012).

³⁸ Arthur E. Wilmarth, Jr., *Beware the Return of the ILC*, American Banker, Aug. 2, 2017, at <https://www.americanbanker.com/opinion/beware-the-return-of-the-ilc>.

³⁹ Press Release, Waters Calls on FDIC to Hold Public Hearing on SoFi’s Application for Bank Charter, at <https://democrats-financialservices.house.gov/news/documentsingle.aspx?DocumentID=400739>.

into very few baskets, and federal regulators will be under great pressure to protect those baskets during future financial and economic disruptions.”⁴⁰

III. Futurist Fintech

Though sober reports from the World Economic Forum, Deloitte, and governmental entities give a good sense of the incrementalist side of fintech, it is important to realize that much of the excitement about the topic of financial technology arises out of a more futuristic perspective. On Twitter, hashtags like #legaltech, #regtech, #insurtech, and #fintech often convene enthusiasts who aspire to revolutionize the financial landscape—or at least to make a good deal of money disrupting existing “trust institutions” (e.g., the intermediaries which help store and transfer financial assets).

Futurist fintech envisions “smart contracts,” which would be executed via some degree of automatic, code-based enforcement.⁴¹ As one article puts it, “Where a smart contract’s conditions depend upon real-world data (e.g., the price of a commodity future at a given time), agreed-upon outside systems, called oracles, can be developed to monitor and verify prices, performance, or other real-world events.”⁴² However, until robotic assessments of physical reality are far less delayed, corroded by a lack of data, and contestable (thanks to the messy complexity of discordant human meanings), the prevalence of totally automated, smart contracts is likely to be limited.

There are many contractual relationships that are too complex and variable, and require too much human judgment, to be reliably coded into software. Code may reflect and in large part implement what the parties intended, but should not *itself* serve as the contract or business agreement among them.

Still, some technologists and lawyers aspire to that subsumption, echoing older movements for financial deregulation.⁴³ The rise of Bitcoin as an alternative currency has

⁴⁰ Arthur E. Wilmarth, Jr., *Beware the Return of the ILC*, American Banker, Aug. 2, 2017, at <https://www.americanbanker.com/opinion/beware-the-return-of-the-ilc>

⁴¹ Joshua Fairfield, *Smart Contracts, Bitcoin Bots, and Consumer Protection*, 71 WASH. & LEE L. REV. ONLINE 35, 38—39 (2014) (“Smart contracts—automated programs that transfer digital assets within the block-chain upon certain triggering conditions—represent a new and interesting form of organizing contractual activity.”).

⁴² Nicolette De Sevres, Bart Chilton & Bradley Cohen, *The Blockchain Revolution, Smart Contracts and Financial Transactions*, 21 NO. 5 CYBERSPACE LAWYER 3, 3 (June 2016). A smart contract is created by encoding the terms of a traditional contract and uploading the smart contract to the blockchain. “Contractual clauses are automatically executed when pre-programmed conditions are satisfied,” and because the transactions are monitored, validated, and enforced by the blockchain, there is no need for a trusted third party, such as an escrow agent. *Id.*

⁴³ DAVID GOLUBIA, *THE POLITICS OF BITCOIN* (2016) (describing parallels between cryptocurrency movement, crypto-anarchist beliefs, and older movements to discredit or dismantle financial regulation and central banking).

sparked an interest in automation of transactions and recordation.⁴⁴ Software can allow distributed computers to transfer information en masse and monitor one another.⁴⁵ Bitcoin is a particular case of using blockchain technology to ensure a durable record of ownership, which is intended to be regulated by code.⁴⁶ Blockchain enthusiasts envision it scaling en masse to serve as a distributed ledger of all manner of transactions.

Given enthusiasm expressed for blockchain at the highest levels of international finance,⁴⁷ governments may soon explore more extensive use of blockchain-based, public ledgers of ownership transactions, such as land records.⁴⁸ Such a digital transition would cut out a fair number of time-consuming steps in current financial processing. Using technology to modernize transactions would seem to be a huge opportunity for saving personnel costs and reducing inconvenience.

Yet there are also reasons for caution. As James Grimmelmann observed in 2005, “software is vulnerable to sudden failure, software is hackable, and software is not

⁴⁴ Joshua Fairfield, *Bitproperty*, 88 S. CAL. L. REV. 805, 805 (May 2015) (“Increased interest in cryptocurrencies has driven the development of a series of technologies for creating public, cryptographically secure ledgers of property interests that do not rely on trust in a specific entity to curate the list.”).

⁴⁵ Michael J. Madison, *Social Software, Groups, and Governance*, 2006 MICH. ST. L. REV. 153, 156 (2006).

⁴⁶ Nicolette De Sevres & Bart Chilton & Bradley Cohen, *The Blockchain Revolution, Smart Contracts and Financial Transactions*, 21 NO. 5 CYBERSPACE LAWYER NL 3, 3 (June 2016). A blockchain is a peer-to-peer network where each computer in the network verifies and records every transaction on the network, where transactions are only recorded on the ledger once the network confirms the validity of the transaction, thus preventing third party manipulation and streamlining the record.

⁴⁷ World Economic Forum, *The future of financial infrastructure: An ambitious look at how blockchain can reshape financial services*, (Aug. 2016) http://www3.weforum.org/docs/WEF_The_future_of_financial_infrastructure.pdf; South African Reserve Bank, *Position Paper on Virtual Currencies*, (Dec. 3, 2014), [https://www.resbank.co.za/RegulationAndSupervision/NationalPaymentSystem\(NPS\)/Legal/Documents/Position%20Paper/Virtual%20Currencies%20Position%20Paper%20%20Final_02of2014.pdf](https://www.resbank.co.za/RegulationAndSupervision/NationalPaymentSystem(NPS)/Legal/Documents/Position%20Paper/Virtual%20Currencies%20Position%20Paper%20%20Final_02of2014.pdf); see also David Mills, et. al., *Distributed ledger technology in payments, clearing and settlement*, Federal Reserve Board (2016) available at <https://www.federalreserve.gov/econresdata/feds/2016/files/2016095pap.pdf>.

⁴⁸ It is at this point unclear whether decentralization via distributed ledger technology would address or exacerbate key problems identified in the Mortgage Electronic Registration Systems, Inc. (MERS) in the wake of the financial crisis. Its implementation of “cloud computing” technology was meant to enable instantaneous transfers of ownership rights within the confines of a centralized database. MERS aspired to remove recording responsibilities from the state to a private entity owned by parties (mortgage lenders) with an interest in ownership disputes. Christopher L. Peterson, *Two Faces: Demystifying the Mortgage Electronic Registration System’s Land Title Theory*, 53 WILLIAM AND MARY LAW REVIEW 111 (2011).

robust.”⁴⁹ No technology has developed that would make the blockchain environment impervious to these problems. Waves of hacking and illicit intrusions have rocked health care institutions,⁵⁰ banks,⁵¹ and even campaigns⁵² and governments.⁵³ While blockchain enthusiasts claim that distributed ledgers help avoid the “honeypot” problem of database centralization (which is an inviting target for hackers), concentration of “mining power” could lead to a 51% attack on even a distributed ledger system. Excessive forking is also a threat to the integrity of such networks.

Moreover, some early adopters of this ideal of self-executing or coded law have experienced troubling and telling failures.⁵⁴ Investors in a “decentralized autonomous organization” (DAO) run on code have already experienced the turbulent and troubling aspects of software-governed legal orders. In early 2016, a hacker managed to take millions of dollars in a fashion unanticipated by the drafters of the code governing the organization. The main organizer of the DAO, Vitalik Buterin had to code a “hard fork” for the organization, which essentially shifted funds from the hacker’s account to an account where the original investors in the project could withdraw their funds.⁵⁵

According to Buterin and other organizers of the DAO, this intervention was a success story: it proved the recoverability of their system. But for advocates of futurist fintech, this was a Pyrrhic victory. The *post hoc* intervention violated the principle of autonomy supposedly at the core of the DAO.⁵⁶ Persons managed the smart contract—not mere code.⁵⁷ In other words, the only way the supposedly smart, incorruptible, automated,

⁴⁹ James Grimmelmann, *Regulation by Software*, 114 YALE L.J. 1719, 1742-44 (2005); see also James Grimmelmann, *Anarchy, Status Updates, and Utopia*, 35 PACE L. REV. 135 (2015) (demonstrating the persistence of governance problems in social software).

⁵⁰ See Jessica Jardine Wilkes, *The Creation of HIPAA Culture: Prioritizing Privacy Paranoia over Patient Care*, 2014 B.Y.U. L. REV. 1213 (2014) (“In 2009, the Office of Civil Rights started recording incidents of PHI breaches and created the “Wall of Shame,” which publicly exposes breaches affecting 500 people or more”).

⁵¹ Paul Merrion, *NY Fed’s role in SWIFT cyber heist prompts House panel data request*, WL 3085306, CQ ROLL CALL 2016. (describing hack of Bangladesh’s central bank).

⁵² Anthony J. Gaughan, *Ramshackle Federalism: America’s Archaic and Dysfunctional Presidential Election System*, 85 FORDHAM L. REV. 1021 (2016). (discussing Russian hackers); Melissa Eddy, *After a Cyberattack, Germany Fears Election Disruption*, N.Y. TIMES, Dec. 8, 2016.

⁵³ Tim McCormack, *The Sony and OPM Double Whammy: International Law and Cyber “Attacks”*, 18 SMU SCI. & TECH. L. REV. 379 (2015).

⁵⁴ Nathaniel Popper, *A Hacking of More Than \$50 Million Dashes Hopes in the World of Virtual Currency*, N.Y. TIMES (June 17, 2016).

⁵⁵ Michael del Castillo, *The Hard Fork: What’s About to Happen to Ethereum and the DAO*, COINDESK July 18, 2016, <http://www.coindesk.com/hard-fork-ethereum-dao/>; Vitalik Buterin, *Hard Fork Completed*, ETHEREUM BLOG (July 20, 2016), <https://blog.ethereum.org/2016/07/20/hard-fork-completed/>.

⁵⁶ Matt Levine, *Blockchain Company’s Smart Contracts Were Dumb*, BLOOMBERG NEWS (June 17, 2016), <https://www.bloomberg.com/view/articles/2016-06-17/blockchain-company-s-smart-contracts-were-dumb>.

⁵⁷ *Id.*

and immutable contract actually protected investors was by *allowing human intervention to change its terms and consequences*. Rather than demonstrating the dispensability of human interventions, the DAO has proved the opposite—the vital necessity of human governance over even extensively coded and computerized forms of human cooperation.

When Primavera De Filippi and Samer Hassan speak of the “incorporation of legal rules into code” and “regulation by code,” culminating in a reliance on code “not only to enforce legal rules, but also to draft and elaborate these rules,” they do not present these phenomena as unalloyed goods.⁵⁸ Rather, they are cautious about the “the prospect of automated legal governance” because it may “reduce the freedoms and autonomy of individuals.”⁵⁹ The answer to these concerns is not to double down on the translation of legal rules into code. Rather, the preservation of human control over financial systems will require an alternative paradigm—a vision of software as a tool to assist persons, rather than a machine replacing them. Nor should policymakers abandon long-standing principles of financial regulation to make way for forms of financial automation that have yet to be proven. There is little evidence that regulation means their “revolutionary promise” would be lost, as it was probably never there in the first place.⁶⁰

IV. Conclusion

This testimony has presented reasons to be cautious about legislative or regulatory efforts to federally preempt state laws now applying to both incrementalist and futuristic fintech. I know that advocates for deregulation will likely argue that imposing a level playing field on fintech and non-fintech firms will harm innovation in the fintech sector. But innovation is not good in itself. The toxic assets at the core of the financial crisis were innovative in many ways, but ultimately posed unacceptable risks.⁶¹ So, too, may the superficially attractive services of many fintech firms.

To be sure, promoters of fintech deregulation may claim that such worries are anecdotal. But many tech firms have only themselves to blame for obscuring what we know about the sector. As I explain in my book *The Black Box Society*, aggressive assertion of trade secrecy claims—both about data collection and use, and the algorithms used to make judgments about us—keep regulators and legislators in the dark about the full range of

⁵⁸ Primavera De Filippi & Samer Hassan, *Blockchain Technology as a Regulatory Technology: From Code is Law to Law is Code*, FIRST MONDAY, 21 (12-5) (2016); <http://firstmonday.org/ojs/index.php/fm/article/view/7113/5657#author>.

⁵⁹ *Id.*

⁶⁰ ADAM GREENFIELD, RADICAL TECHNOLOGIES 303 (2017) (“the inventors of the blockchain overtly intended to erode statism and central administration. Virtually everywhere, decision algorithms are touted to us on the promise that they will permanently displace human subjectivity and bias. And yet in every instance we find that these ambitions are flouted, as the technologies that were supposed to enact them are captured...by existing concentrations of power.”).

⁶¹ JENNIFER TAUB, OTHER PEOPLE’S HOUSES (2015).

risks in fintech.⁶² If there is any message I can deliver to the committee today, it is to empower agencies like CFPB and the OFR, and to expand their funding, as they try to come to grips with a rapidly financial landscape.

Data gathering is important, because nearly every story of technologized “financial inclusion” can be countered with other stories of exclusion, via digital redlining. As Cathy O’Neil’s book *Weapons of Math Destruction* shows, consumers often are in the dark about what new algorithms are judging them, and how they can respond if they think they’ve been treated unfairly.⁶³ Regulators need to understand more fully what these firms are doing, and how they are performing. Moreover, as the recent Equifax hack shows, concentration of information in almost any firm creates great risks to consumers. Improving financial cybersecurity should be an essential goal in fintech policy.⁶⁴ I applaud the GAO for highlighting security issues in its report, and Senator Jack Reed for proposing forward-thinking legislation on this front.

We should not have faith that accelerated deregulation will free the financial sector to solve important social problems. The value proposition of some fintechs merely points out larger problems in existing credit provision that could be solved by more direct action. For example, if fintechs can make a hefty profit by refinancing student debts owed to the U.S. government, perhaps that is less an indication of fintechs’ business prowess, than it is evidence that the government is overcharging students for loans.⁶⁵ If consumers are desperate for marketplace lending to cover next month’s utility bills, maybe we need to ensure work pays more fairly, rather than plying them with digital loans. I am confident

⁶² FRANK PASQUALE, *THE BLACK BOX SOCIETY* (2015).

⁶³ CATHY O’NEIL, *WEAPONS OF MATH DESTRUCTION* (2016).



⁶⁴ Kristin Johnson, *Managing Cyber Risks*, 50 Ga. L. Rev. 547 (2016), at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2847234 (discussing SEC cyber-risk management disclosure obligations); Kristin Johnson and Steven Ramirez, *Sustainability: A New Guiding Principle for Financial Market Regulation*, 11 U. ST. THOMAS L. J. 386 (2015).



⁶⁵ Michael Simkovic, *The Knowledge Tax*, U. Chi. L. Rev. (2015), at <http://chicagounbound.uchicago.edu/uclrev/vol82/iss4/4/>; Marc Nerlove, *Some Problems in the Use of Income-contingent Loans for the Finance of Higher Education*, 83 J. POL. ECON. 157, 160, 180 (1975). When private sector refinancers can “cherry pick” or “cream skim” the most creditworthy borrowers from a federal credit program, that risk selection eventually leaves the government dependent on repayment by the worst credit risks. That erodes the sustainability of the federal loan program—and its borrower protections, like income based repayment. See Frank Pasquale, *Democratizing Higher Education: Defending & Extending Income Based Repayment Programs*, 28 LOY. CONSUMER L. REV. 1 (2015), at <http://lawcommons.luc.edu/lclr/vol28/iss1/2>, for more on the politics of public finance accounting and the role of private lenders in undermining the perceived and actual sustainability of federal credit programs.

that a system of postal banking would do far more than the fintech sector to deliver financial inclusion to the millions of Americans without adequate access to deposit accounts.⁶⁶

In conclusion: Fintech should not be an excuse for more regulatory arbitrage. We need far more information about how fintech firms are gathering and processing data. And we should be wary about the ability of technology alone to solve much larger social problems of financial inclusion, opportunity, and fair, non-discriminatory credit provision.



⁶⁶ MEHRSA BARADARAN, *HOW THE OTHER HALF BANKS* (2015). Over 25% of US households are unbanked or underbanked. FDIC, *FDIC National Survey of Unbanked and Underbanked Households* (2016).



 **Josh Constine**  @JoshConstine · Nov 22 ▼
Some tech billionaire, please buy out the local ISP(s) where FCC chairman Ajit Pai lives and give him 14.4k dial-up speeds for killing net neutrality
🗨️ 24 ↻ 409 ❤️ 1.3K ✉️



 **Matthew Prince**  @eastdakota ▼
Replying to @JoshConstine
I could do this in a different, but equally effective, way.
7:20 PM · Nov 22, 2017


110 Retweets **428** Likes

🗨️ ↻ ❤️ ✉️

 **Josh Constine**  @JoshConstine · Nov 22 ▼
Replying to @eastdakota
Please, do go on...
🗨️ 1 ↻ 2 ❤️ 32 ✉️

 **Matthew Prince**  @eastdakota · Nov 22 ▼
sent note to our GC to see if we can without breaking any laws.
🗨️ 16 ↻ 12 ❤️ 164 ✉️

 **Josh Constine**  @JoshConstine · Nov 22 ▼
Thanks for doing your part and please let me know how this goes. If he wants an unregulated internet, he might not like what he gets
🗨️ 2 ↻ 1 ❤️ 51 ✉️

 **Matthew Prince**  @eastdakota · Nov 22 ▼
indeed.
🗨️ 3 ↻ ❤️ 51 ✉️

epic.org

Electronic Privacy Information Center
1718 Connecticut Avenue NW, Suite 200
Washington, DC 20009, USA

+1 202 483 1140
+1 202 483 1248
@EPICPrivacy
<https://epic.org>

November 28, 2017

The Honorable Robert Latta, Chair
The Honorable Marsha Blackburn, Chair
The Honorable Janice Schakowsky, Ranking Member
The Honorable Michael Doyle, Ranking Member
U.S. House Committee on Energy and Commerce
Subcommittee on Digital Commerce & Consumer Protection
Subcommittee on Communications and Technology
2125 Rayburn House Office Building
Washington, D.C. 20515

Dear Chairs Latta and Blackburn and Ranking Members Schakowsky and Doyle:

We write to you regarding the "Algorithms: How Companies' Decisions About Data and Content Impact Consumers" hearing.¹ EPIC is a public interest research center established in 1994 to focus public attention on emerging privacy and civil liberties issues.² EPIC has promoted "Algorithmic Transparency" for many years.³

Democratic governance is built on principles of procedural fairness and transparency. And accountability is key to decision making. We must know the basis of decisions, whether right or wrong. But as decisions are automated, and organizations increasingly delegate decisionmaking to techniques they do not fully understand, processes become more opaque and less accountable. It is therefore imperative that algorithmic process be open, provable, and accountable. Arguments that algorithmic transparency is impossible or "too complex" are not reassuring.

It is becoming increasingly clear that Congress must regulate AI to ensure accountability and transparency:

- Algorithms are often used to make adverse decisions about people. Algorithms deny people educational opportunities, employment, housing, insurance, and credit.⁴ Many of

¹ *Algorithms: How Companies' Decisions About Data and Content Impact Consumers*, 115th Cong. (2017), H. Comm. on Energy & Commerce, Subcomm. on Digital Commerce and Consumer Protection and Subcomm. on Communications and Technology, <https://energycommerce.house.gov/hearings/algorithms-companies-decisions-data-content-impact-consumers/> (Nov. 29, 2017).

² EPIC, *About EPIC*, <https://epic.org/epic/about.html>.

³ EPIC, *Algorithmic Transparency*, <https://epic.org/algorithmic-transparency/>.

⁴ Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 Wash. L. Rev. 1 (2014).

EPIC Statement
House E&C Committee

1

Algorithmic Transparency
November 28, 2017

Defend Privacy. Support EPIC.

these decisions are entirely opaque, leaving individuals to wonder whether the decisions were accurate, fair, or even about them.

- Secret algorithms are deployed in the criminal justice system to assess forensic evidence, determine sentences, to even decide guilt or innocence.⁵ Several states use proprietary commercial systems, not subject to open government laws, to determine guilt or innocence. The Model Penal Code recommends the implementation of recidivism-based actuarial instruments in sentencing guidelines.⁶ But these systems, which defendants have no way to challenge are racially biased, unaccountable, and unreliable for forecasting violent crime.⁷
- Algorithms are used for social control. China's Communist Party is deploying a "social credit" system that assigns to each person government-determined favorability rating. "Infractions such as fare cheating, jaywalking, and violating family-planning rules" would affect a person's rating.⁸ Low ratings are also assigned to those who frequent disfavored web sites or socialize with others who have low ratings. Citizens with low ratings will have trouble getting loans or government services. Citizens with high ratings, assigned by the government, receive preferential treatment across a wide range of programs and activities.
- In the United States, U.S. Customs and Border Protection has used secret analytic tools to assign "risk assessments" to U.S. travelers.⁹ These risk assessments, assigned by the U.S. government to U.S. citizens, raise fundamental questions about government accountability, due process, and fairness. They may also be taking us closer to the Chinese system of social control through AI.

In a recent consumer complaint to the Federal Trade Commission, EPIC challenged the secret scoring of young athletes.¹⁰ As EPIC's complaint regarding the Universal Tennis Rating system makes clear, the "UTR score defines the status of young athletes in all tennis related activity; impacts opportunities for scholarship, education and employment; and may in the future provide the basis for 'social scoring' and government rating of citizens."¹¹ As we explained to

⁵ *EPIC v. DOJ (Criminal Justice Algorithms)*, EPIC, <https://epic.org/foia/doj/criminal-justice-algorithms/>; *Algorithms in the Criminal Justice System*, EPIC, <https://epic.org/algorithmic-transparency/crim-justice/>.

⁶ Model Penal Code: Sentencing §6B.09 (Am. Law. Inst., Tentative Draft No. 2, 2011).

⁷ Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

⁸ Josh Chin & Gillian Wong, *China's New Tool for Social Control: A Credit Rating for Everything*, Wall Street J., Nov. 28, 2016, <http://www.wsj.com/articles/chinas-new-tool-for-social-control-a-credit-rating-for-everything-1480351590>

⁹ *EPIC v. CBP (Analytical Framework for Intelligence)*, EPIC, <https://epic.org/foia/dhs/cbp/afi/>.

¹⁰ EPIC, *EPIC Asks FTC to Stop System for Secret Scoring of Young Athletes* (May 17, 2017), <https://epic.org/2017/05/epic-asks-ftc-to-stop-system-f.html>; See also Shanya Possess, *Privacy Group Challenges Secret Tennis Scoring System*, Law360, May 17, 2017, <https://www.law360.com/articles/925379>; Lexology, *EPIC Takes a Swing at Youth Tennis Ratings*, June 1, 2017, <https://www.lexology.com/library/detail.aspx?g=604e3321-dfc8-4f46-9afc-abd47c5a5179>

¹¹ EPIC Complaint to Federal Trade Commission, In re Universal Tennis at 1 (May 17, 2017).

the FTC, “EPIC seeks to ensure that all rating systems concerning individuals are open, transparent and accountable.”¹²

In *re Universal Tennis*, EPIC urged the FTC to (1) Initiate an investigation of the collection, use, and disclosure of children’s personal information by Universal Tennis; (2) Halt Universal Tennis’s scoring of children without parental consent; (3) Require that Universal Tennis make public the algorithm and other techniques that produce the UTR; (4) Require that Universal Tennis establish formal procedures for rectification of inaccurate, incomplete, and outdated scoring procedures; and (5) Provide such other relief as the Commission finds necessary and appropriate.¹³

“Algorithmic Transparency” must be a fundamental principle for consumer protection.¹⁴ The phrase has both literal and figurative dimensions. In the literal sense, it is often necessary to determine the precise factors that contribute to a decision. If, for example, a government agency or private company considers a factor such as race, gender, or religion to produce an adverse decision, then the decision-making process should be subject to scrutiny and the relevant factors identified.

Some have argued that algorithmic transparency is simply impossible, given the complexity and fluidity of modern processes. But if that is true, there must be some way to recapture the purpose of transparency without simply relying on testing inputs and outputs. We have seen recently that it is almost trivial to design programs that evade testing.¹⁵ And central to the science and innovation is the provability of results.

Europeans have long had a right to access “the logic of the processing” concerning their personal information.¹⁶ That principle is reflected in the U.S. in the publication of the FICO score, which for many years remained a black box for consumers, establishing credit worthiness without providing any information about the basis of score.¹⁷

The continued deployment of AI-based systems raises profound issues for democratic countries. As Professor Frank Pasquale has said:

Black box services are often wondrous to behold, but our black box society has become dangerously unstable, unfair, and unproductive. Neither New York quants nor California engineers can deliver a sound economy or a secure society. Those

¹² *Id.*

¹³ *Id.* at 13.

¹⁴ *At UNESCO, Rotenberg Argues for Algorithmic Transparency*, EPIC (Dec. 8, 2015), <https://epic.org/2015/12/at-unesco-epics-rotenberg-argu.html>.

¹⁵ Jack Ewing, *In '06 Slide Show, a Lesson in How VW Could Cheat*, N.Y. Times, Apr. 27, 2016, at A1.

¹⁶ Directive 95/46/EC—The Data Protection Directive, art 15 (1), 1995,

<http://www.dataprotection.ie/docs/EU-Directive-95-46-EC--Chapter-2/93.htm>.

¹⁷ Hadley Malcom, *Banks Compete on Free Credit Score Offers*, USA Today, Jan. 25, 2015, <http://www.usatoday.com/story/money/2015/01/25/banks-free-credit-scores/22011803/>.

are the tasks of a citizenry, which can perform its job only as well as it understands the stakes.¹⁸

We ask that this Statement from EPIC be entered in the hearing record. We look forward to working with you on these issues of vital importance to the American public.

Sincerely,

/s/ Marc Rotenberg
Marc Rotenberg
EPIC President

/s/ Caitriona Fitzgerald
Caitriona Fitzgerald
EPIC Policy Director

¹⁸ Frank Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information* 218 (Harvard University Press 2015).

GREG WALDEN, OREGON
CHAIRMAN

FRANK PALLONE, JR., NEW JERSEY
RANKING MEMBER

ONE HUNDRED FIFTEENTH CONGRESS
Congress of the United States
House of Representatives
COMMITTEE ON ENERGY AND COMMERCE
2125 RAYBURN HOUSE OFFICE BUILDING
WASHINGTON, DC 20515-6115
Majority (202) 225-2927
Minority (202) 225-3641
December 15, 2017

Dr. Omri Ben-Shahar
Leo and Eileen Professor of Law
Kearney Director, Coase-Sandor Institute for Law and Economics
University of Chicago Law School
1111 East 60th Street
Chicago, IL 60615

Dear Dr. Ben-Shahar:

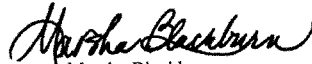
Thank you for appearing before the Subcommittee on Communications and Technology and the Subcommittee on Digital Commerce and Consumer Protection on Wednesday, November 29, 2017, to testify at the joint hearing entitled "Algorithms: How Companies' Decisions About Data and Content Impact Consumers."

Pursuant to the Rules of the Committee on Energy and Commerce, the hearing record remains open for ten business days to permit Members to submit additional questions for the record, which are attached. The format of your responses to these questions should be as follows: (1) the name of the Member whose question you are addressing, (2) the complete text of the question you are addressing in bold, and (3) your answer to that question in plain text.

To facilitate the printing of the hearing record, please respond to these questions with a transmittal letter by the close of business on Tuesday, January 9, 2018. Your responses should be mailed to Evan Viau, Legislative Clerk, Committee on Energy and Commerce, 2125 Rayburn House Office Building, Washington, DC 20515 and e-mailed to Evan.Viau@mail.house.gov.

Thank you again for your time and effort preparing and delivering testimony before the Subcommittees.

Sincerely,



Marsha Blackburn
Chairman
Subcommittee on Communications
and Technology



Robert E. Latta
Chairman
Subcommittee on Digital Commerce
and Consumer Protection

cc: The Honorable Michael F. Doyle, Ranking Member, Subcommittee on Communications and Technology

The Honorable Janice D. Schakowsky, Ranking Member, Subcommittee on Digital Commerce and Consumer Protection

Attachment



1111 East 60th Street | Chicago, Illinois 60637
 phone 773-702-2087 | fax 773-702-0730
 e-mail omri@uchicago.edu
 home.uchicago.edu/omri

Omri Ben-Shahar
Leo Herzel Professor in Law
Director, Coase-Sandor Institute for Law and Economics

Response to Additional Questions for the Record
“Algorithms:
How Companies’ Decisions About Data and Content Impact Consumers”
 by

Professor Omri Ben-Shahar, University of Chicago

Submitted to

Committee on Energy and Commerce

Subcommittee on Communications and Technology

Subcommittee on Digital Commerce and Consumer Protection

The Honorable Robert E. Latta

1. Do algorithmic technologies and processes pose unique challenges for a disclosure and consent model of data collection

Response: Algorithmic technologies pose heightened, and in my view insurmountable, challenges for disclosure and consent model of data collection. Evidence shows that disclosures have failed to produce “informed” consent in areas that involve less complex subject matter. In the area of algorithmic technologies, the underlying information that needs to be communicated to consumers in order to accomplish meaningful consent is large and complex. Consumers need to be told what information is collected and used by algorithms, how it shared by companies, and how it is secured. Because consumers visit many apps and websites daily, the amount of information necessary makes any model of meaningful consumer choice entirely unrealistic.

The Honorable Gregg Harper

1. What limits does the First Amendment place on the government requiring or preventing disclosure of certain information?

Response: I am not a constitutional lawyer and cannot provide authoritative response on this matter.

2. In your testimony, you make it clear that you believe disclosure regimes are ineffectual. What are the alternatives?

Response: The primary regulatory alternative is a regime that prohibits some types of algorithmic information collections and uses altogether. It reduces risk, but also benefit. I do not recommend this, other than in extremely vulnerable areas.

The critical requirement for the design of any good alternative regulatory solution is to diagnose the problem that the solution seeks to address. In the area of data privacy, the questions “what is the consumer injury” and “what is the risk to society” have not been answered convincingly, and until then we should not devise solutions.

It is clear, however, algorithmic data collection is leading to market power and concentration. Large companies are getting larger *because* of the data they have. They are able to tailor more personalized and satisfying experience for their customers and gain further edge on their competitors. Accordingly, the best way to regulate these markets is to guarantee thriving entry and competition.

The Honorable Michael C. Burgess

1. In your testimony, you state that mandatory disclosure requirements don’t work because people agree to terms without reading them in order to access the content or application. Does access to content or applications have to be all or nothing?
 - a. Is it possible for a consumer to agree to limited or no collection of their information and still gain access to their desired content?

Response: It is possible to design partial collection options, that grant consumers partial access to websites and applications. It is challenging, however, to create a *tool* that enables consumers to choose smartly between the options. Companies would, naturally, try to direct consumers towards “checking the boxes” that best serve the companies’ commercial interests. It would require sophistication, knowledge, and savvy among consumers to be able to pick the best “limited” option. Evidence in this area and in many other consumer markets suggest such efforts would most likely fail.

2. Algorithms have vast positive potential and capacity from self-driving cars to medicine and public health, including helping to find cures for diseases. Is there a way to help consumers understand algorithms as they apply to all sectors that may affect our daily lives?

Response: No. The complexity of the issues involve defeat such educational goal

The Honorable Adam Kinzinger

- 1.a. How can government and the content providers work together to balance free speech and our national interest?
- 1.b. Can data be considered speech? Can data collection be covered by the First Amendment?

Response: I am not an expert on these matters and regrettably cannot provide authoritative responses to these important questions.

GREG WALDEN, OREGON
CHAIRMAN

FRANK PALLONE, JR., NEW JERSEY
RANKING MEMBER

ONE HUNDRED FIFTEENTH CONGRESS
Congress of the United States
House of Representatives
COMMITTEE ON ENERGY AND COMMERCE
2125 RAYBURN HOUSE OFFICE BUILDING
WASHINGTON, DC 20515-6115
Majority (202) 225-2927
Minority (202) 225-3641
December 15, 2017

Ms. Kate Klonick
Resident Fellow
Information Society Project
Yale Law School
110 4th Avenue, #4A
Brooklyn, NY 11217

Dear Ms. Klonick:

Thank you for appearing before the Subcommittee on Communications and Technology and the Subcommittee on Digital Commerce and Consumer Protection on Wednesday, November 29, 2017, to testify at the joint hearing entitled "Algorithms: How Companies' Decisions About Data and Content Impact Consumers."

Pursuant to the Rules of the Committee on Energy and Commerce, the hearing record remains open for ten business days to permit Members to submit additional questions for the record, which are attached. The format of your responses to these questions should be as follows: (1) the name of the Member whose question you are addressing, (2) the complete text of the question you are addressing in bold, and (3) your answer to that question in plain text.

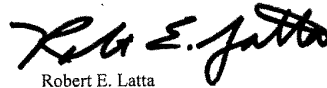
To facilitate the printing of the hearing record, please respond to these questions with a transmittal letter by the close of business on Tuesday, January 9, 2018. Your responses should be mailed to Evan Viau, Legislative Clerk, Committee on Energy and Commerce, 2125 Rayburn House Office Building, Washington, DC 20515 and e-mailed to Evan.Viau@mail.house.gov.

Thank you again for your time and effort preparing and delivering testimony before the Subcommittees.

Sincerely,



Marsha Blackburn
Chairman
Subcommittee on Communications
and Technology



Robert E. Latta
Chairman
Subcommittee on Digital Commerce
and Consumer Protection

cc: The Honorable Michael F. Doyle, Ranking Member, Subcommittee on Communications and Technology

The Honorable Janice D. Schakowsky, Ranking Member, Subcommittee on Digital Commerce and Consumer Protection

Attachment

[Ms. Klonick did not answer submitted questions for the record by the time of printing.]

Additional Questions for the RecordThe Honorable Robert E. Latta

1. Can data be considered speech, and can data collection be covered by the First Amendment?

The Honorable Adam Kinzinger

1. Content providers have a remarkable challenge with respect to deeming content as “inappropriate”. There are a multitude of problems that may arise in some part of the process. First, the sheer volume of content out there is staggering. On YouTube alone, approximately 400 hours of content is uploaded every minute. A social media platform or some other content provider may enable consumers to “rate” or “like or dislike” content, which could then trigger a flag and initiate a review process, but that process is subjective and would almost certainly involve human error. A provider may be able to use algorithms to facilitate a review process, but there is just no way to write an algorithm that would catch everything. A provider may use human screeners, but again, the sheer volume of content is so overwhelming that it seems impossible to be able to review all content. Lastly, and perhaps most importantly, even if a provider uses a combination of these tools and processes, the term itself—“inappropriate”—is oftentimes subjective. There are scenarios in which a video may be posted that offends those with certain political views. That video may even be offensive to the vast majority of Americans. One might think this sort of scenario would normally be left to one of our branches of government to decide, but the content providers are doing the jobs themselves.
 - a. In the consumer based media platforms available today, are content providers able to police their content?
 - b. Do you believe that it is acceptable for them to regulate themselves?
 - c. If content providers have enough trouble reviewing and regulating content, I am skeptical that government would be able to do it any better. But do you feel differently? What would such a process look like?
2. Because many platforms’ algorithms are proprietary, they are often compared to as a “black box”, where user information is collected with an unclear purpose. What can you tell us about the impact of content shaping and delivery of targeted ads on consumers’ right to Constitutionally protected speech? What’s the nexus?
 - a. Is there a difference between average users versus public figures?
 - b. Why do public figures have different standards?
 - c. What is the legal basis for these differences?

GREG WALDEN, OREGON
CHAIRMAN

FRANK PALLONE, JR., NEW JERSEY
RANKING MEMBER

ONE HUNDRED FIFTEENTH CONGRESS
Congress of the United States
House of Representatives
COMMITTEE ON ENERGY AND COMMERCE
2125 RAYBURN HOUSE OFFICE BUILDING
WASHINGTON, DC 20515-6115
Majority (202) 225-2927
Minority (202) 225-3641
December 15, 2017

Ms. Laura Moy
Deputy Director
Georgetown Law Center on Privacy and Technology
600 New Jersey Avenue, N.W.
Washington, DC 20001

Dear Ms. Moy:

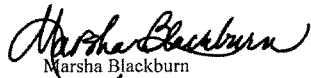
Thank you for appearing before the Subcommittee on Communications and Technology and the Subcommittee on Digital Commerce and Consumer Protection on Wednesday, November 29, 2017, to testify at the joint hearing entitled "Algorithms: How Companies' Decisions About Data and Content Impact Consumers."

Pursuant to the Rules of the Committee on Energy and Commerce, the hearing record remains open for ten business days to permit Members to submit additional questions for the record, which are attached. The format of your responses to these questions should be as follows: (1) the name of the Member whose question you are addressing, (2) the complete text of the question you are addressing in bold, and (3) your answer to that question in plain text.

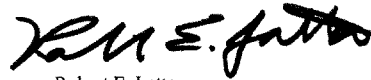
To facilitate the printing of the hearing record, please respond to these questions with a transmittal letter by the close of business on Tuesday, January 9, 2018. Your responses should be mailed to Evan Viau, Legislative Clerk, Committee on Energy and Commerce, 2125 Rayburn House Office Building, Washington, DC 20515 and e-mailed to Evan.Viau@mail.house.gov.

Thank you again for your time and effort preparing and delivering testimony before the Subcommittees.

Sincerely,



Marsha Blackburn
Chairman
Subcommittee on Communications
and Technology



Robert E. Latta
Chairman
Subcommittee on Digital Commerce
and Consumer Protection

cc: The Honorable Michael F. Doyle, Ranking Member, Subcommittee on Communications and Technology

The Honorable Janice D. Schakowsky, Ranking Member, Subcommittee on Digital Commerce and Consumer Protection

Attachment



January 9, 2018

The Honorable Marsha Blackburn, Chairman
The Honorable Jan Schakowsky, Ranking Member
Committee on Energy & Commerce
Subcommittee on Communications and Technology

The Honorable Bob Latta, Chairman
The Honorable Michael Doyle, Ranking Member
Committee on Energy & Commerce
Subcommittee on Digital Commerce and Consumer Protection

Dear Chairman Blackburn, Chairman Latta, Ranking Member Doyle, and
Ranking Member Schakowsky:

Thank you again for inviting me to testify in the November 29, 2017
hearing on "Algorithms: How Companies' Decisions About Data and
Content Impact Consumers." Below please find my responses to the
additional questions for the record.

Sincerely,



Laura M. Moy

Attachment

1. **As you know, online bias—racial or otherwise—is not limited to websites. If left unchecked, broadband providers also have the power to make decisions that skew our national dialogue and harm free speech. These choices can be driven by economic considerations or even forced through political pressure.**

To shield against these problems, FCC adopted strong net neutrality protections in 2015. These rules ensure that consumers can decide for themselves what they see online. But for some reason, the Trump FCC is now planning to wipe out these critical safeguards. Do you believe that FCC's proposal to eliminate net neutrality could lead to more bias online?

The FCC's elimination of net neutrality rules could indeed lead to more bias online. Net neutrality rules would have prevented Internet service providers (ISPs) from turning Internet access into a pay-for-play service, where content may be blocked, throttled, or prioritized based on ISPs' own judgments about what they want Internet users to see.

Without the rules, ISPs may choose to prioritize some viewpoints expressed online over others based on which online speakers can afford to pay for increased access to Internet users. This would lead to greater bias online.

2. **Many Congressional Republicans have cheered on FCC's efforts to destroy net neutrality. They argue that eliminating these important protections will somehow benefit consumers. Republicans are actually making some of the very same arguments that they used earlier this year to justify the elimination of FCC's privacy rules using the Congressional Review Act. I am curious whether those arguments ended up panning out the way Republicans claimed at the time. I know you have been following the privacy debate for a long time. From your perspective, are consumers better off as a result of the Republicans' elimination of FCC's privacy rules?**

Consumers are not better off as a result of the 2017 Congressional Review Act resolution that eliminated federal broadband privacy regulations. At the time Congress voted to eliminate those consumer protections, Energy and Commerce Committee Chairman Greg Walden said, "Once these rules are reversed, the FCC can again work effectively with the FTC to ensure that our privacy framework allows the internet to flourish while truly protecting

consumers.”¹ Communications and Technology Subcommittee Chairman Marsha Blackburn said, “Today’s action takes us one step closer to restoring the FTC’s role as America’s expert agency on privacy.”

But since broadband privacy rules were eliminated, the FCC and FTC have not worked to protect the privacy of broadband subscribers. In December the agencies released a memorandum of understanding outlining their approach to protecting consumers and the public interest online.² The word “privacy” does not appear in the memorandum.

A year ago, consumers were the beneficiaries of strong prospective rules that clearly outlined what ISPs could and could not do with their customers’ private information. Today, no such rules exist. If the recent repeal of the 2015 reclassification order stands, in the future the FTC may take enforcement action against ISPs that violate their own privacy policies. But this does not provide consumers with any confidence that an ISP with which they have no choice to share highly private information—information about where they go and what they do online—will not be used in ways that consumers find invasive and offensive.

3. **The Federal Communications Commission (FCC) Chairman has often relied on a solitary justification for eliminating FCC’s net neutrality protections. He claims that broadband providers’ investments have decreased because of net neutrality. I’ve seen data that was submitted to FCC that shows the opposite is true.**
 - a. **What is the difference between what broadband providers told FCC regarding their investment under the net neutrality rules versus what they are telling their investors and the SEC?**

¹ House Energy & Commerce Committee, *Press Release: House Advances Resolution Rolling Back FCC’s Flawed ISP Privacy Rules* (Mar. 28, 2017), <https://energycommerce.house.gov/news/press-release/house-advances-resolution-rolling-back-fcc-s-flawed-isp-privacy-rules/>.

² Fed. Trade Comm’n and Fed. Communications Comm’n, *Restoring Internet Freedom: FCC-FTC Memorandum of Understanding*, https://www.ftc.gov/system/files/documents/cooperation_agreements/fcc_ftc_mou_internet_freedom_order_1214_final_0.pdf.

ISPs told the FCC that complying with net neutrality rules would harm their ability to invest in the network. On the contrary, ISPs' capital investments *increased* over the two years following the FCC's February 2015 Open Internet Order in February 2015. This is detailed in depth in a robust study of ISP-industry companies' own reports to their investors and to the Securities and Exchange Commission, published last year by Free Press.³ According to that report, "The total capital investment by publicly traded ISPs was 5 percent higher during the two-year period following the FCC's Open Internet vote than it was in the two years prior to the vote."⁴ In addition, "None of the firms that saw declines [in capital spending] attributed them to any FCC action."⁵

b. Does the Draft FCC Order account for the increase in investment by other companies that use the internet to deliver their services?

The FCC Order also does not account for capital investments made by non-ISP-industry companies following the 2015 Open Internet Order. Companies that use the Internet to deliver their services appear to have increased their investments under strong net neutrality rules. This increase in investments may even be attributable to net neutrality, because net neutrality rules increased edge providers' confidence that the Internet would remain a neutral playing field.

This is also detailed in the study published by Free Press, which explains, "Capital investments in edge-computing industry sectors grew dramatically in the wake of the FCC's restoration of its authority to protect these nondiscriminatory telecom services."⁶

³ S. Derek Turner, Free Press, *It's Working: How the Internet Access and Online Video Markets Are Thriving in the Title II Era*, <https://www.freepress.net/sites/default/files/resources/internet-access-and-online-video-markets-are-thriving-in-title-II-era.pdf>.

⁴ *Id.* at 4.

⁵ *Id.*

⁶ *Id.* at 7.

GREG WALDEN, OREGON
CHAIRMAN

FRANK PALLONE, JR., NEW JERSEY
RANKING MEMBER

ONE HUNDRED FIFTEENTH CONGRESS
Congress of the United States
House of Representatives
COMMITTEE ON ENERGY AND COMMERCE
2125 RAYBURN HOUSE OFFICE BUILDING
WASHINGTON, DC 20515-6115
Majority (202) 225-2927
Minority (202) 225-3641
December 15, 2017

Dr. Catherine Tucker
Sloan Distinguished Professor of Management Science
MIT Sloan School of Management
100 Main Street, E62-536
Cambridge, MA 02142

Dear Dr. Tucker:


Thank you for appearing before the Subcommittee on Communications and Technology and the Subcommittee on Digital Commerce and Consumer Protection on Wednesday, November 29, 2017, to testify at the joint hearing entitled "Algorithms: How Companies' Decisions About Data and Content Impact Consumers."


Pursuant to the Rules of the Committee on Energy and Commerce, the hearing record remains open for ten business days to permit Members to submit additional questions for the record, which are attached. The format of your responses to these questions should be as follows: (1) the name of the Member whose question you are addressing, (2) the complete text of the question you are addressing in bold, and (3) your answer to that question in plain text.

To facilitate the printing of the hearing record, please respond to these questions with a transmittal letter by the close of business on Tuesday, January 9, 2018. Your responses should be mailed to Evan Viau, Legislative Clerk, Committee on Energy and Commerce, 2125 Rayburn House Office Building, Washington, DC 20515 and e-mailed to Evan.Viau@mail.house.gov.

Thank you again for your time and effort preparing and delivering testimony before the Subcommittees.

Sincerely,


Marsha Blackburn
Chairman
Subcommittee on Communications
and Technology


Robert E. Latta
Chairman
Subcommittee on Digital Commerce
and Consumer Protection

cc: The Honorable Michael F. Doyle, Ranking Member, Subcommittee on Communications and Technology

The Honorable Janice D. Schakowsky, Ranking Member, Subcommittee on Digital Commerce and Consumer Protection

Attachment

Massachusetts Institute of
Technology
Sloan School of Management
Catherine Tucker
*Sloan Distinguished Professor of
Management
Professor of Marketing
Chair MIT Sloan PhD Program*



Management Science Area
Phone: (617) 252-1499
E-mail: cetucker@mit.edu
cetucker.scripts.mit.edu/

TO: Marsha Blackburn, Chairman of the Subcommittee on Communications and
Technology, Robert E. Latta, Chairman of the Subcommittee on Digital Commerce and
Consumer Protection
FROM: Catherine Tucker
DATE: January 2018

**Responses to additional questions arising from Testimony on 'Algorithms: How
Companies' Decisions about Data and Content Impact Consumers'**

Dear Congresswoman Blackburn and Congressman Latta,

In response to your letter dated December 15, 2017, I would like to offer additional
answers to the supplementary questions submitted by members of Congress.

The Honorable Robert E. Latta

1. *Your research has demonstrated that economic factors
can provide for different outcomes than perhaps even the creators of the relevant
algorithms might have intended. Assumptions might be built into the algorithm about
what does or does not count, but the output might not be reliable or intended due to
unforeseen factors not built in the model. Can a regulatory technique be fashioned to
solve these types of problems, or is it another way of saying that highly complex systems
at this juncture in time will tend to produce unexpected results?*

I think you are exactly right to hone in on the question of what regulatory technique may
be most appropriate at this time of uncertainty and transition.

My research so far has been mainly focused on the question of whether "algorithmic
transparency" is sufficient or necessary as a regulatory regime for algorithms. My
research suggests it is neither sufficient nor particularly helpful as a policy emphasis. The
reason algorithmic transparency is not sufficient is that there are many cases where the
data that the algorithm feeds on, not the algorithm itself, is what causes bias (or at least
the appearance of bias). Just looking at the algorithmic code (supposing that was even
viable or possible) would not allow regulators to identify instances of bias in such cases.
The reason it is not helpful is that "hard coding" of bias or discrimination into a code is
rare, from my experience in talking to many technology companies. It is simply not the
case that programmers add lines to their code where they instruct the algorithm to treat

groups differently on the basis of race or gender. Instead, a more appropriate area of concern is a focus on the complex interactions of algorithms with data and human behavior, with a particular focus on studying outcomes relative to their non-algorithmic counterfactual.

2. *What impact have prescriptive regulations, such as those promulgated under the Children's Online Privacy Protection Act, had on investment and creation of content for children?*

We have a new paper on the question of children's privacy in mobile applications. It has not been released yet, but I hope to have a copy up online by February and will email your staff to ensure the committee has access. This paper focuses on mobile applications on smartphones that are targeted at the under-5s. We find two main things:

- 1) A surprising number of applications targeted at toddlers or preschoolers collect highly personalized data (including precise location data).
- 2) Many of these applications are developed by developers outside the US, and in particular the most intrusive applications often originate from developers based in Asia and countries like Ukraine.

One interpretation of these results is that potentially developers within the US are not developing apps for children because of concerns over legal compliance, and that absence has attracted foreign developers who are not constrained by any concerns for children's privacy. Consequently, we may inadvertently have a situation where our children's privacy laws may have led to worse privacy practices in the apps on the market.

The Honorable Gregg Harper

1. Can you tell us some of the considerations consumers make when deciding to exchange private information for services, and the degree to which existing disclosure rules factor into those decisions?

This has been a great deal of research into this question over the last few years, which I probably cannot do justice to except for saying that this seems to be very context dependent. My own research highlights that more sophisticated or technologically-savvy consumers often are unwilling to share personal data, unless it is framed in terms of an economic exchange whereby there is some gain (even if it is negligible) in doing so. My research also suggests that the effect of disclosure rules depends on the extent to which they are accompanied by a parallel sense of control for the consumer. Simply receiving information about the potential risks of data disclosure can be off-putting to consumers, unless they are offered (even a slight) sense of control at the same time.

The Honorable Michael C. Burgess

1. *In your testimony, you state that "algorithms may appear biased." In your research, how does algorithmic bias manifest itself - how do you measure issues of bias or fairness?*

My research has been focused on algorithms where the outcome may appear biased, but the bias reflects market outcomes, rather than human bias as such. The way we measured this apparent bias was by seeing whether or not women or men were more likely to see ads for jobs in Science, Technology, Engineering and Math. We found that women were less likely to see such ads, not because of any direct bias or because the algorithm predicted that women were less likely to respond to the ad. Instead, it was because women are such a desirable demographic that they cost more to advertise to, and the algorithm, in its attempt to save the advertiser money, showed fewer ads to expensive female eyeballs.

This particular example, I think, is useful in illustrating how hard it is to say whether an algorithm is "biased" or "fair" or "discriminatory." Instead, we have a well-meaning algorithm trying to be cost-effective, which inadvertently leads to an outcome where women see fewer job ads in a way we may find as a society undesirable. I would hesitate to call this 'bias,' but instead think of it as an example of the occasional inadvertent consequences of well-intentioned algorithms leading to outcomes that are less than desirable.

2. *If an algorithm tends to produce results that were not intended by its creator, what is the likelihood that fact will be discovered and corrected?*

I think this will be context dependent. If I were to speculate, my guess is that firms who are developing specialized algorithms for obviously sensitive areas (such as predictive policing, predictive sentencing, enhancing hiring decisions) will be more likely to conduct audits and ensure that their results are not inadvertently distorted.

My concerns would instead focus on firms that are developing algorithms whose client base is broad enough that they may not be aware that there will be particular cases or situations where algorithmic bias may matter. One example of this is the advertising industry. On the whole, we don't really care as a society who sees a particular shoe ad. On the other hand, there are isolated cases where we do care who sees advertising - for example, we might care if discriminated-against racial groups were more likely to see ads from predatory lenders, and we might care if women are less likely to see ads for high-paying jobs than men.

The Honorable Adam Kinzinger

1. Given that companies tend to have extensive and rather transparent privacy policies, does more disclosure tend to make consumers more reluctant to use a particular service or site?

My research suggests that in general more disclosure can have a chilling effect. This is partially because consumers can find it off-putting, but also because complying with disclosure requirements can impose costs on firms, meaning they may not offer that particular service. The exception to this is when disclosure is accompanied by a parallel sense of control for consumers. In such instances, consumers are encouraged to use a technology. So for example, if a company has a set of disclosures but also communicates to a consumer that they retain control or ownership of their own data, that can increase the chance of a customer using a service or website.

I hope these responses are helpful. Please let me know if you have any concerns or if there is anything I can clarify.

Yours sincerely,

A large black rectangular redaction box covering the signature of Catherine Tucker.

Catherine Tucker

GREG WALDEN, OREGON
CHAIRMAN

FRANK PALLONE, JR., NEW JERSEY
RANKING MEMBER

ONE HUNDRED FIFTEENTH CONGRESS
Congress of the United States
House of Representatives
COMMITTEE ON ENERGY AND COMMERCE
2125 RAYBURN HOUSE OFFICE BUILDING
WASHINGTON, DC 20515-6115
Majority (202) 225-2927
Minority (202) 225-3641
December 15, 2017

Mr. Frank Pasquale
Professor of Law
University of Maryland Frances King Carey School of Law
500 West Baltimore Street
Baltimore, MD 21201

Dear Mr. Pasquale:


Thank you for appearing before the Subcommittee on Communications and Technology and the Subcommittee on Digital Commerce and Consumer Protection on Wednesday, November 29, 2017, to testify at the joint hearing entitled "Algorithms: How Companies' Decisions About Data and Content Impact Consumers."

Pursuant to the Rules of the Committee on Energy and Commerce, the hearing record remains open for ten business days to permit Members to submit additional questions for the record, which are attached. The format of your responses to these questions should be as follows: (1) the name of the Member whose question you are addressing, (2) the complete text of the question you are addressing in bold, and (3) your answer to that question in plain text.

To facilitate the printing of the hearing record, please respond to these questions with a transmittal letter by the close of business on Tuesday, January 9, 2018. Your responses should be mailed to Evan Viau, Legislative Clerk, Committee on Energy and Commerce, 2125 Rayburn House Office Building, Washington, DC 20515 and e-mailed to Evan.Viau@mail.house.gov.

Thank you again for your time and effort preparing and delivering testimony before the Subcommittees.

Sincerely,


Marsha Blackburn
Chairman
Subcommittee on Communications
and Technology


Robert E. Latta
Chairman
Subcommittee on Digital Commerce
and Consumer Protection

cc: The Honorable Michael F. Doyle, Ranking Member, Subcommittee on Communications and Technology

The Honorable Janice D. Schakowsky, Ranking Member, Subcommittee on Digital Commerce and Consumer Protection

Attachment

Additional Questions for the Record

Research for the following responses was prepared by Jennifer Smith, Beasley Fellow at the University of Maryland, at the request of Frank Pasquale, Professor of Law at the University of Maryland.

Questions from the Honorable Janice D. Schakowsky

1. Last year, a ProPublica investigation found that Facebook allowed housing ads targeting users by race, religion, disability, nationality, and other protected traits. Federal law forbids housing ads that include “any preference, limitation, or discrimination” based on such traits.

In response, Facebook built an automated system intended to prevent biased ads. But ProPublica published a follow-up report in November 2017 showing that the exact same biased ads were approved by Facebook’s new automated system.

- a. Is a fully automated system sufficient to prevent biased or discriminatory advertising? If not, what else if necessary?
- b. Some critics have called housing ads that exclude certain ZIP codes or neighborhoods a modern day form of redlining. Location-based housing ads can be helpful, but are you concerned about their potential for bias? Should these ads be subject to additional review?

Short answers:

- a) No, automation cannot prevent such bias, and may actually accelerate it. Human review with involvement from affected groups is important to preventing such discrimination.
- b) Location-based ads should be subject to additional review, thanks to the rise of data-based proxies for race and other protected characteristics.

General Information to support these points:

- On November 29, 2017 *ProPublica* reported, “Facebook said it would temporarily stop advertisers from being able to exclude viewers by race while it studies the use of its ads targeting system.”¹
 - This article was published the same day as the Committee Hearing.

This raises concerns about the legal standard for housing discrimination, and potential secondary liability for platforms.

- Disparate Impact
 - Key background here is the U.S. Department of Housing and Urban Development’s (HUD) 2013 rule on disparate impact liability, and the Supreme Court’s 2015 decision in

¹ Julia Angwin, *Facebook to Temporarily Block Advertisers from Excluding Audiences by Race*, PROPUBLICA (Nov. 29, 2017, 2:00 PM), <https://www.propublica.org/article/facebook-to-temporarily-block-advertisers-from-excluding-audiences-by-race>.

*Texas Department of Housing and Community Affairs v. The Inclusive Communities Project, Inc.*²

- The Supreme Court found “disparate-impact claims are cognizable under the Fair Housing Act” in the 2015 case *Texas Department of Housing and Community Affairs v. The Inclusive Communities Project, Inc.*³
 - Although the Court formally found “disparate impact claims are cognizable under the Fair Housing Act,” the Court also set a standard of proof that many commentators view as narrowing disparate impact liability.⁴
- In 2013 HUD issued *Implementation of the Fair Housing Act’s Discriminatory Effects Standard; Final Rule*⁵ which “established a three-part burden shifting test.”⁶
 - “First, the plaintiff must demonstrate that the challenged practice caused or predictably will cause a discriminatory effect. Then, the burden shifts to the defendant to prove that the challenged practice is necessary to achieve one or more ‘substantial, legitimate, nondiscriminatory interests.’ If the defendant satisfies that burden, then, the plaintiff must prove that the substantial, legitimate, nondiscriminatory interest could be accomplished through a practice that has a less discriminatory effect. The defendant will be able to prevail if it can show that the substantial, legitimate, nondiscriminatory interest cannot be achieved through a practice that has any less discriminatory effect.”⁷
- To date it remains unclear how the Court’s decision in *Texas Department of Housing and HUD’s* 2013 rule will impact HUD enforcement of Fair Housing Act claims⁸

² Tex. Dep’t of Hous. & Cmty. Affairs v. Inclusive Cmty. Project, Inc., 135 S. Ct. 2507 (2015) (opinion available through HUD’s website at <https://www.hud.gov/sites/documents/131371BSACUNITEDSTATES.PDF>).

³ Tex. Dep’t of Hous. & Cmty. Affairs v. Inclusive Cmty. Project, Inc., 135 S. Ct. 2507 (2015) (opinion available through HUD’s website at <https://www.hud.gov/sites/documents/131371BSACUNITEDSTATES.PDF>).

⁴ See MICHAEL W. SKOJEC & MICHAEL P. CIANFINCHI, NAT’L MULTIFAMILY HOUSING COUNCIL & NAT’L APARTMENT ASSN., RECENT HUD ACTIONS REGARDING DISPARATE IMPACT (Apr. 2017), <https://www.naahq.org/sites/default/files/naa-documents/government-affairs/protected/recent-hud-actions-regarding-disparate-impact.pdf>; Paul Hancock, *Symposium: The Supreme Court Recognizes but Limits Disparate Impact in its Fair Housing Act Decision*, SCOTUSBLOG (Jun. 26, 2015, 8:58 AM), <http://www.scotusblog.com/2015/06/paul-hancock-fha/>.

⁵ 24 C.F.R. Part 100, Vol. 78, No. 3, 11460 (Feb. 15, 2013) (rule available through HUD’s website at <https://www.hud.gov/sites/documents/DISCRIMINATORYEFFECTRULE.PDF>).

⁶ MICHAEL W. SKOJEC & MICHAEL P. CIANFINCHI, NAT’L MULTIFAMILY HOUSING COUNCIL & NAT’L APARTMENT ASSN., RECENT HUD ACTIONS REGARDING DISPARATE IMPACT (Apr. 2017), <https://www.naahq.org/sites/default/files/naa-documents/government-affairs/protected/recent-hud-actions-regarding-disparate-impact.pdf>.

⁷ MICHAEL W. SKOJEC & MICHAEL P. CIANFINCHI, NAT’L MULTIFAMILY HOUSING COUNCIL & NAT’L APARTMENT ASSN., RECENT HUD ACTIONS REGARDING DISPARATE IMPACT (Apr. 2017), <https://www.naahq.org/sites/default/files/naa-documents/government-affairs/protected/recent-hud-actions-regarding-disparate-impact.pdf> (citations omitted).

⁸ See MICHAEL W. SKOJEC & MICHAEL P. CIANFINCHI, NAT’L MULTIFAMILY HOUSING COUNCIL & NAT’L APARTMENT ASSN., RECENT HUD ACTIONS REGARDING DISPARATE IMPACT (Apr. 2017), <https://www.naahq.org/sites/default/files/naa-documents/government-affairs/protected/recent-hud-actions-regarding-disparate-impact.pdf>; Eric Epstein et al.,

- Current HUD Secretary Ben Carson's 2015 *Washington Times* Op-Ed criticizing Obama administration HUD actions as "government-engineered attempts to legislate racial equality" may point to HUD not strongly enforcing disparate impact claims⁹
- Secondary Liability
 - A September 20, 2006 Memorandum from the U.S. Department of Housing and Urban Development states:
 - "The prohibition applies to all advertising media, including newspapers, magazines, television, radio, and the Internet. Just as the Department has found newspapers in violation of the Fair Housing act for publishing discriminatory classifieds, the Department has also concluded that it is illegal for Web sites to publish discriminatory advertisements."¹⁰
- There is a current case against Facebook stemming from Facebook ads' discriminatory practices:
 - Following *ProPublica's* 2016 report on Facebook ads' discriminatory practices a complaint was filed in the U.S. District Court for the Northern District of California seeking "declaratory relief, injunctive relief, penalties, and monetary damages under the Fair Housing Act . . . and Title VII of the Civil Rights Act of 1964 to redress discrimination" based on Facebook Ads' inclusion of the "Exclude People" button.¹¹
 - Facebook filed a motion to dismiss in April 2017, arguing Section 230 of the Communications Decency Act bars the plaintiffs' claims against Facebook since the ads were originated by third parties.
 - In making their argument Facebook cited:
 - Fair Housing Council of San Fernando Valley v. Roommates.com, LLC, 512 F.3d 1157 (9th Cir. 2008)
 - Perfect 10, Inc. v. CCBill LLC, 488 F.3d 1102 (9th Cir. 2007)

The U.S Supreme Court's Decision in Texas Department of Housing & Community Affairs v. Inclusive Communities Project, Inc., DORSEY & WHITNEY LLP PUBLICATIONS (June 30, 2015), <https://www.dorsey.com/newsresources/publications/client-alerts/2015/06/the-us-supreme-courts-decision-in-texas-departm>.

⁹ Ben S. Carson, Op-Ed, *Experimenting with Failed Socialism Again*, WASH. TIMES (July 23, 2015), <https://www.washingtontimes.com/news/2015/jul/23/ben-carson-obamas-housing-rules-try-to-accomplish/>.

¹⁰ Memorandum from Bryan Green, Deputy Assisnt Sec'y for Enforcement Programs, Fair Housing Act Application to Internet Advertising (Sept. 20, 2006), <https://www.hud.gov/sites/documents/INTERNETADVERTMEMO.PDF>. See also Written Statement by Kim Kendrick, Ass't Sec. of Fair Housing & Equal Opportunity, before the U.S. House of Rep., fair Housing Issues in the Gulf Coast in the Aftermath of Hurricanes Katrina and Rita (Feb. 28, 2006), <https://archives.hud.gov/testimony/2006/test022806.cfm> (stating "HUD has received and is investigating complaints alleging that some Internet sites have carried advertisements offering housing to evacuees, but only if they were of the right race or religion, or have no children. The Fair Housing Act makes it unlawful to publish discriminatory statements in connection with the sale or rental of housing.").

¹¹ *Onuoha v. Facebook, Inc.*, No. 5:16-cv-06440-EJD (N.D. Cal. 2016) (complaint is available at <http://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?article=2320&context=historical>). The case is also cited as *Mobley v. Facebook, Inc.*

- Currently, Facebook’s motion to dismiss is administratively terminated while the parties are in mediation.¹²

This raises a question of whether HUD could investigate Facebook for its discriminatory ads (reported on ProPublica)? What would be the maximum penalties? Could states get involved?

- HUD investigation and penalties
 - The Fair Housing Act states it is unlawful to “make, print, or publish, or cause to be made, printed, or published any notice, statement, or advertisement, with respect to the sale or rental of a dwelling that indicates any preference, limitation, or discrimination based on race, color, religion, sex, handicap, familial status, or national origin, or an intention to make any such preference, limitation, or discrimination.”¹³
 - An individual who believes his or her rights under the Fair Housing Act have been violated can file a complaint with HUD and HUD will investigate and attempt to conciliate the issue¹⁴
 - HUD may also charge a party with violation of the Fair Housing Act and assess a civil penalty in addition to any actual damages and/or attorneys’ fees and costs¹⁵
 - The current civil penalty amounts are: “a maximum civil penalty of \$19,787 for his or her first violation of the Fair Housing Act. Respondents who had violated the Fair Housing Act in the previous 5 years could be fined a maximum of \$49,467, and respondents who had violated the Act two or more times in the previous 7 years could be fined a maximum of \$98,935.”¹⁶
 - According to *ProPublica*, following *ProPublica*’s 2016 report HUD was “in discussions’ with Facebook to address what it termed ‘serious concerns’ about the social network’s advertising practices.”¹⁷

¹² See Joint Post-Mediation Status Report (Oct. 30, 2017) (report is available at <https://www.leagle.com/decision/infdc020171101c16>).

¹³ 42 U.S.C. 3604(c) (2015).

¹⁴ 42 U.S.C. 3610 (2015); *Fair Housing – It’s Your Right*, HUD, https://www.hud.gov/program_offices/fair_housing_equal_opp/online-complaint (last visited Jan. 5, 2018).

¹⁵ For information on HUD enforcement activity since 2004, see *Fair Housing Act Enforcement Activity*, HUD, https://www.hud.gov/program_offices/fair_housing_equal_opp/enforcement (last visited Jan. 5, 2018).

¹⁶ Inflation Catch-Up Adjustments of Civil Monetary Penalty Amounts, 81 Fed. Reg. 38,931 (June 15, 2016) (to be codified at 24 C.F.R. pts. 28, 30, 87, 180, and 3282); Jeff Dillman, *HUD Increases Civil Penalty Amounts for Fair Housing Violations*, Fair Hous. Proj. (June 16, 2016), <http://www.fairhousingnc.org/2016/hud-increases-civil-penalty-amounts-fair-housing-violations/>.

¹⁷ Stephen Engelberg, *HUD has Serious Concerns about Facebook’s Ethnic Targeting*, PROPUBLICA (Nov. 7, 2016, 4:27 PM), <https://www.propublica.org/article/hud-has-serious-concerns-about-facebooks-ethnic-targeting> (quoting HUD spokeswoman Heather Fluitt). See also Teke Wiggin, *HUD Discussing ‘Serious Concerns’ with Facebook Over Ad Targeting*, Inman (Nov. 3, 2016), <https://www.inman.com/2016/11/03/hud-discussing-serious-concerns-with-facebook-over-ad-targeting/>.

- The New York Times “sought to dismiss the suit on First Amendment grounds, arguing that the ads were created by advertisers and that the newspaper ‘merely published the advertisements as submitted.’”²³
- The New York Times settled the case in 1993, agreeing to pay \$150,000 in damages and donating \$300,000 worth of advertising to the New York Open Housing Center. Further, the Times implemented a “policy requiring that pictures of people in housing advertisements be representative of the racial makeup in the metropolitan area.”²⁴
- It appears most media organizations and online platforms are being held accountable in the “court of public opinion” and in investigative articles like those of *ProPublica* for aiding and abetting discrimination, but not by government agencies and courts.

2. Some platforms rely on users to report advertising and content that violate a platform’s standards. Is reporting by users sufficient, or do platforms need a proactive system so that such content and advertising are never approved and published in the first place?

To prevent very bad outcomes, platforms need a proactive system to avoid certain very troubling content from becoming widely disseminated. Fortunately, they are now taking this responsibility more seriously, but more needs to be done.

- Ads and content being vetted by humans
 - Recent commentary advocates a combination of machine and human interventions to identify content that violates a platform’s standards.
 - A recent article in *Quartz* states “[t]he solution will likely require a combination of machines and humans, where the machines flag phrases that appear to be offensive, and humans decide whether those phrases amount to hate speech.”²⁵
 - Frank Pasquale’s article “The Automated Public Sphere” also suggests this need for human review.
 - There appears to be a shift away from platforms relying only on users and monitoring technology to flag content, toward the hiring of more humans to monitor content.
 - For example, the *Financial Times* reported in December 2017: “Both YouTube and Facebook have previously relied heavily on users reporting inappropriate content and technology designed to root it out automatically. The companies have changed tack and are now investing in more people, although they continue to hope that improvements in machine learning will make the removal of content more efficient.”

²³ William Glaberson, *Times Adopts a New Policy in Advertising for Housing*, N.Y. TIMES (Aug. 14, 1993), <http://www.nytimes.com/1993/08/14/nyregion/times-adopts-a-new-policy-in-advertising-for-housing.html>.

²⁴ William Glaberson, *Times Adopts a New Policy in Advertising for Housing*, N.Y. TIMES (Aug. 14, 1993), <http://www.nytimes.com/1993/08/14/nyregion/times-adopts-a-new-policy-in-advertising-for-housing.html>; *New York Times Settles Ad Case Worth \$450,000*, NAT’L FAIR HOUSING ADVOC. ONLINE (1993), <https://fairhousing.com/news-archive/advocate/1993/new-york-times-settles-ad-case-worth-450000>.

²⁵ Keith Collins, *Facebook and Google Need Humans, Not Just Algorithms, to Filter Out Hate Speech*, QUARTZ (Sept. 17, 2017), <https://qz.com/1075499/facebook-and-google-need-humans-not-just-computers-to-filter-out-hate-speech/>.

- Google employs “ads quality raters,” temporary workers who watch videos on YouTube to “identify and flag offensive material to build the trove of data [Google’s] AI will learn from.”²⁶
 - The information is not used to remove videos, rather the information is used to combat the issue of paid ads being shown with videos promoting violence, hate speech, and terrorism. Recently, companies such as Walmart and PepsiCo stopped advertising on YouTube due to the uncertainty of whether their ads would be shown with offensive videos.²⁷
 - In December 2017, Google announced it would hire more human reviewers to review content on YouTube in response to advertiser concerns as well as parents concerned about reports of violent content being targeted at children.²⁸
 - In the fall of 2017, Facebook announced it would “begin subjecting ads targeted based on social issues, politics, religion, and ethnicity to human review.”²⁹
 - Sheryl Sandberg, Facebook’s Chief Operating Officer, posted on her Facebook account that Facebook was “adding more human review and oversight to our automated processes.”³⁰
- Content-flagging systems that automatically pull down copyrighted content and offensive/inappropriate content show that a fully automated system can lead to many false positives and other difficulties. Human judgment remains necessary now, and in the foreseeable future.
 - Background on pulling down copyrighted content
 - For general information see:
 - YouTube Help’s *How Content ID Works*³¹
 - The Electronic Frontier Foundation’s *A Guide to YouTube Removals*³²
 - Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473 (2016),

²⁶ Davey Alba, *The Hidden Laborers Training AI to Keep Ads Off Hateful YouTube Videos*, WIRED (Apr. 21, 2017, 2:08 PM), <https://www.wired.com/2017/04/zerochaos-google-ads-quality-raters/>.

²⁷ Davey Alba, *The Hidden Laborers Training AI to Keep Ads Off Hateful YouTube Videos*, WIRED (Apr. 21, 2017, 2:08 PM), <https://www.wired.com/2017/04/zerochaos-google-ads-quality-raters/>.

²⁸ Hannah Kuchler, *YouTube Hires Moderators to Root Out Inappropriate Videos*, FIN. TIMES (Dec. 5, 2017), <https://www.ft.com/content/080d1dd4-d92c-11e7-a039-c64b1c09b482>.

²⁹ Kevin Tran, *Humans will Vet Political Ads on Facebook*, BUS. INSIDER (Oct. 10, 2017, 9:39 AM), <http://www.businessinsider.com/humans-will-vet-political-ads-on-facebook-2017-10>. See also Todd Spangler, *Facebook Pledges to Hire 1,000 More Ad Reviewers Amid Russian Political Scandal*, VARIETY (Oct. 2, 2017, 8:55 PM), <http://variety.com/2017/digital/news/facebook-to-hire-1000-ad-reviewers-russian-political-scandal-1202577789/>.

³⁰ Sheryl Sandberg, FACEBOOK (Sept. 20, 2017), <https://www.facebook.com/sheryl/posts/10159255449515177>. See also David Ingram, *Facebook to Add More Human Review to Ad System – COO Sandberg*, REUTERS (Sept. 20, 2017, 9:34 PM), <https://www.reuters.com/article/legal-us-facebook-advertising/facebook-to-add-more-human-review-to-ad-system-coo-sandberg-idUSKCN1BV2X5>.

³¹ YOUTUBE HELP, HOW CONTENT ID WORKS, <https://support.google.com/youtube/answer/2797370?hl=en> (last visited Jan. 4, 2017).

³² A GUIDE TO YOUTUBE REMOVALS, ELECTRONIC FRONTIER FOUND., <https://www.eff.org/issues/intellectual-property/guide-to-youtube-removals#content-id> (last visited Jan. 4, 2017).

<https://law.stanford.edu/wp-content/uploads/2016/10/Accountability-in-Algorithmic-Copyright-Enforcement.pdf>.

- Frank Pasquale, *The Black Box Society*, Chapter 3.³³
- Facebook has recently taken steps to improve flagging unauthorized content:
 - In July 2017 Facebook bought the intellectual property tracking company Source3³⁴
 - Facebook announced in October 2017 it would integrate “Rights Manager with services from three third-party providers” in order “to make it easier for content owners to police the social platform for unauthorized and pirated videos.”³⁵
- Pull down inappropriate/offensive content
 - In November 2017, Facebook stated “99% of the ISIS and Al Qaeda-related terror content we remove from Facebook is content we detect before anyone in our community has flagged it to us, and in some cases, before it goes live on the site. We do this primarily through the use of automated systems like photo and video matching and text-based machine learning. Once we are aware of a piece of terror content, we remove 83% of subsequently uploaded copies within one hour of upload.”³⁶

3. Algorithms can be manipulated to promote content that is dangerous. For example, conspiracy theories opposing vaccines are sometimes disproportionately promoted on social media platforms. What kinds of content and engagement do social media platforms’ algorithms favor when “deciding” what to put in our feeds? Is there a risk that they may disproportionately favor sensationalist content that may not be true?

There is a bias toward content that increases engagement—and often this is very sensationalistic or even untrue content.³⁷ As David Golumbia has argued:

³³ Frank Pasquale, *The Black Box Society* (Cambridge: Harvard University Press, 2015); Frank Pasquale, *Dominant search engines: an essential cultural & political facility*, in *The Next Digital Decade* (2010), at http://digitalcommons.law.umaryland.edu/cgi/viewcontent.cgi?article=2363&context=fac_pubs.

³⁴ Todd Spangler, *Facebook Buys Startup Source3 to Get Better at Catching Pirated Content*, VARIETY (July 25, 2017, 8:12 AM), <http://variety.com/2017/digital/news/facebook-acquires-source3-piracy-1202505740/>.

³⁵ Todd Spangler, *Facebook Connects Video Copyright-Flagging System to Third-Party Tools*, VARIETY (Oct. 3, 2017, 7:00 AM), <http://variety.com/2017/digital/news/facebook-rights-manager-copyright-videos-third-party-1202578122/>.

³⁶ Monika Bickert, *Hard Questions: Are We Winning the War on Terrorism Online?*, FACEBOOK NEWSROOM (Nov. 28, 2017), <https://newsroom.fb.com/news/2017/11/hard-questions-are-we-winning-the-war-on-terrorism-online/>. See also Hannah Kuchler, *Facebook Says It Can Quickly Remove Most Content From Terrorist Groups*, FIN. TIMES (Nov. 28, 2017), <https://www.ft.com/content/a2ff9a9e-3230-371a-90e4-339a0ffc61c0>; Jana J. Pruet, *Facebook Says It Deletes 99 Percent of ISIS and Al Qaeda Content Before It’s Flagged*, THE BLAZE (Nov. 29, 2017, 4:30 PM), <http://www.theblaze.com/news/2017/11/29/facebook-says-it-deletes-99-percent-of-isis-and-al-qaeda-content-before-its-flagged>.

³⁷ Frank Pasquale, *The Automated Public Sphere*, at (2017)

Social media too easily bypasses the rational or at least reasonable parts of our minds, on which a democratic public sphere depends. It speaks instead to the emotional, reactive, quick-fix parts of us, that are satisfied by images and clicks that look pleasing, that feed our egos, and that make us think we are heroic. But too often these feelings come at the expense of the deep thinking, planning, and interaction that democratic politics are built from. This doesn't mean reasoned debate can't happen online; of course it can and does. It means that there is a strong tendency—what media and technology researchers call an “affordance”—away from dispassionate debate and toward strong emotions.³⁸

There is also evidence that certain manipulators can pollute or otherwise influence feeds. For every change a social media platform makes to an algorithm there are multiple sources instructing users (usually focused on marketing) on how to get around or “beat” the change. See, e.g.:

- Betsy McLeod, *How to Beat Facebook's News Feed Algorithm*, BLUE CORONA MARKETING BLOG (Aug. 9, 2017), <https://www.bluecorona.com/blog/facebook-news-feed-algorithm-tips>.
- Jenn Chen, *9 Tips to Improve Organic Growth with the Facebook Algorithm*, SPROUT BLOG (July 19, 2017), <https://sproutsocial.com/insights/facebook-algorithm/>.
- Gabriele Boland, *How Brands Can Adapt to Social Media Algorithms*, NEWSWHIP (Apr. 14, 2016), <http://www.newswhip.com/2016/04/brands-can-adapt-social-media-algorithms/>.
- Christina Newberry, *The Twitter Algorithm: What You Need to Know to Boost Organic Reach*, HOOTSUITE BLOG (May 15, 2017), <https://blog.hootsuite.com/twitter-algorithm/>.

The following sources give more information on the nature of the problem:

Citation	Description/Quotes
Samuel Albanie, Hillary Shakespeare & Tom Gunter, <i>Unknowable Manipulators: Social Network Curator Algorithms</i> , 30TH CONFERENCE ON NEURAL INFO. PROCESSING SYS. (2016), http://www.robots.ox.ac.uk/~albanie/publications/albanie16manipulators.pdf .	Authors discuss how algorithms used to engage users can also learn to manipulate users.
JANNA ANDERSON & LEE RAINIE, PEW RES. CTR., THE FUTURE OF TRUTH AND MISINFORMATION ONLINE (Oct. 2017), http://www.pewinternet.org/2017/10/19/the-future-of-truth-and-misinformation-online/ .	Report on survey of over 1000 “technologists, scholars, practitioners, strategic thinkers and others” about the “online information environment.” “51% chose the option that the information environment will not improve, and 49% said the information environment will improve.”

³⁸ David Golumbia, *Social Media Has Hijacked our Brains Threatens Global Democracy*, at https://motherboard.vice.com/en_us/article/bjy7ez/social-media-threatens-global-democracy.

<p>FACEBOOK: NEWS FEED, https://newsfeed.fb.com/welcome-to-news-feed?lang=en (last visited Jan. 5, 2018).</p> <p>For updates on changes to the Facebook News Feed, see <i>News Feed FYI</i>, Facebook Newsroom, https://newsroom.fb.com/news/category/news-feed-fyi/ (last visited Jan. 5, 2018).</p>	<p>Information on Facebooks News Feed.</p>
<p>Wael Ghonim & Jake Rashbass, <i>It's Time to End the Secrecy and Opacity of Social Media</i>, WASH. POST (Oct. 31, 2017), https://www.washingtonpost.com/news/democracy-post/wp/2017/10/31/its-time-to-end-the-secrecy-and-opacity-of-social-media/?utm_term=.5e2b2c180484.</p>	<p>Authors advocate for “far more transparency of the outputs produced by [social media] algorithms so we can create an effective accountability mechanism” and a “standardized public interest API.”</p>
<p>Andrew Hutchinson, <i>How Twitter's Feed Algorithm Works – As Explained by Twitter</i>, Social Media Today (May 11, 2017), https://www.socialmediatoday.com/social-networks/how-twitters-feed-algorithm-works-explained-twitter.</p>	<p>Author describes how Twitter’s algorithm works, including how tweets are ranked.</p>
<p>DILIP KRISHNA, NANCY ALBINSON & YANG CHU, DELOITTE, <i>MANAGING ALGORITHMIC RISKS (2017)</i>, https://www2.deloitte.com/us/en/pages/risk/articles/algorithmic-machine-learning-risk-management.html.</p>	<p>Authors examine “algorithmic risks” and offer advice to organizations/businesses on how best to manage those risks.</p>
<p>Nicolas Koumchatzky & Arton Andryeyev, <i>Using Deep Learning at Scale in Twitter's Timeline</i>, TWITTER: ENGINEERING (May 9, 2017), https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-twitters-timelines.html.</p>	<p>Describing how Twitter’s “ranking algorithm is powered by deep neural networks.”</p>
<p>Will Oremus, <i>Twitter's New Order</i>, SLATE (Mar. 5, 2017, 8:00 PM), http://www.slate.com/articles/technology/cover_story/2017/03/twitter_s_timeline_algorithm_and_its_effect_on_us_explained.html.</p>	<p>Author examines Twitter’s “algorithmic timeline.”</p> <p>“But you can’t see more of some kinds of tweets without seeing less of others, and the hidden consequences of that equation could affect us all. As it gradually tightens the loops in Twitter’s social fabric, the algorithm risks further insulating its users from people whose viewpoints run counter to their own—a phenomenon, already rampant on Facebook, that has contributed to the polarization of the American electorate and the Balkanization of its media.”</p>

	<p>“Twitter, in other words, is no longer a social network, at least by its own reckoning. It’s a real-time, personalized news service. And since there are no human editors, it falls to Twitter’s algorithm to determine which tweets will lead the news each time you open it.”</p> <p>“Yet if ever-greater personalization is the answer to Twitter’s business woes, it’s unlikely to be the answer to the woes of a media ecosystem in which all news has become “fake news” to someone.”</p>
<p>LEE RAINIE & JANNA ANDERSON, PEW RES. CTR., CODE-DEPENDENT: PROS AND CONS OF THE ALGORITHM AGE (Feb. 2017), http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/.</p>	<p>Report on survey of over 1300 “technology experts, scholars, corporate practitioners and government leaders” on “current attitudes about the potential impacts of algorithms in the next decade.”</p> <p>“The non-scientific canvassing found that 38% of these particular respondents predicted that the positive impacts of algorithms will outweigh negatives for individuals and society in general, while 37% said negatives will outweigh positives; 25% said the overall impact of algorithms will be about 50-50, positive-negative.”</p>
<p>Kai Shu, Amy Silva, Suhang Wang, Jiliang Tang & Huan Liu, <i>Fake News Detection on Social Media: A Data Mining Perspective</i>, ARXIV (Sept. 3, 2017), https://arxiv.org/pdf/1708.01967v3.pdf.</p>	<p>Paper discusses the dissemination and risks related to fake news online as well as an overview of current detection strategies and research.</p>
<p>Tom Wheeler, <i>Using “Public Interest Algorithms” to Tackle the Problems Created by Social Media Algorithms</i>, BROOKINGS: TECHTANK (Nov. 1, 2017), https://www.brookings.edu/blog/techtank/2017/11/01/using-public-interest-algorithms-to-tackle-the-problems-created-by-social-media-algorithms/.</p>	<p>Author proposes use of public interest algorithms “to monitor and report on the effects of social media algorithms.”</p> <p>“[A] public interest algorithm can provide awareness of and access to the information behind any posting. Such sunlight will not only expose any propaganda, but also will help independent evaluation of the veracity of the information being delivered.”</p> <p>“That problem is how the software algorithms that determine what you see on social media prioritize revenue over veracity.” (emphasis added).</p>

The Honorable Frank Pallone, Jr.

1. In your testimony for this hearing, you discussed ways that racial bias can leak into the content we see online. I appreciate your work in this area pointing out a problem that the Congressional Black Caucus has also been working hard to address. I am concerned that systematic bias in our technology could cause disproportionate harms to minority communities.

Fortunately, on our Committee, Congressman Butterfield has led the fight along with Congressman Rush and Congresswoman Clarke to tackle this issue head on. Off Committee, Congressman Cleaver, Congressman Ellison, and Congresswoman Lee have also taken the problem straight to the tech companies, forcing them to confront their role creating this widespread problem. Are there ways that technology companies can better wring bias out from our systems?

Yes, and this is a critical problem. Groups like the Electronic Privacy Information Center, AI Now, Algorithm Watch, Data & Society, and Upturn have worked on this problem for years. They have generated many key reports which the Committee should consult.³⁹ The following resources give further information:

Citation	Description/Quotes
<p>DEEPMIND ETHICS & SOCIETY, https://deepmind.com/applied/deepmind-ethics-society/ (last visited Jan. 8, 2018).</p> <p>See also Verity Harding & Sean Legassick, <i>Why We Launched DeepMind Ethics & Society</i>, DeepMind (Oct. 3, 2017), https://deepmind.com/blog/why-we-launched-deepmind-ethics-society/.</p>	<p>“Technology is not value neutral, and technologists must take responsibility for the ethical and social impact of their work.”</p>
<p>Jen Heazlewood, <i>Combatting Unconscious Bias in Design</i>, R/GA BY DESIGN (Feb. 2, 2017), https://rgabydesign.com/combating-unconscious-bias-in-design-ac5940232fb7.</p>	<p>“The result of the actions by designers quickly encroaches on that of machines, and as we progress further into the world of machine learning and artificial intelligence we need to ensure that pre-existing models and shortcuts are not designed into the technology. A problem with the evolution of these systems is that algorithms are being created with the inventors’ unconscious biases: Once systems are created, the test subjects are often internal subjects or recruits with similar backgrounds to the creators, therefore the voice or learning program becomes more receptive to that uniform group.”</p>
<p>Matt Reynolds, <i>Bias Test to Prevent Algorithms Discriminating Unfairly</i>, NEW SCIENTIST (Mar. 29,</p>	

³⁹ See also Giovanni Comandè, *Regulating Algorithms’ Regulation: First Ethico-Legal Principles*, in *Transparent Data Mining for Big and Small Data* (edited by Tania Cerquitelli, Daniele Quercia, and Frank Pasquale, 2016).

<p>2017), https://www.newscientist.com/article/mg23431195-300-bias-test-to-prevent-algorithms-discriminating-unfairly/.</p>	
<p>Jackie Snow, <i>New Research Aim to Solve the Problem of AI Bias in "Black Box" Algorithms</i>, MIT TECH. REV. (Nov. 7, 2017), https://www.technologyreview.com/s/609338/new-research-aims-to-solve-the-problem-of-ai-bias-in-black-box-algorithms/.</p>	<p>Author discusses recent research and proposal to combat algorithmic bias.</p> <p>See Sarah Tan et al. paper cited below.</p>
<p>Matthias Spielkamp, <i>Inspecting Algorithms for Bias</i>, MIT TECH. REV. (June 12, 2017), https://www.technologyreview.com/s/607955/inspecting-algorithms-for-bias/.</p>	<p>"Democratic societies should be working now to determine how much transparency they expect from ADM systems. Do we need new regulations of the software to ensure it can be properly inspected? Lawmakers, judges, and the public should have a say in which measures of fairness get prioritized by algorithms. But if the algorithms don't actually reflect these value judgments, who will be held accountable?"</p>
<p>Sarah Tan, Rich Caruana, Giles Hooker & Yin Lou, <i>Detecting Bias in Black-Box Models Using Transparent Model Distillation</i>, ARXIV (Nov. 18, 2017), https://arxiv.org/pdf/1710.06169.pdf.</p>	<p>Authors discuss and "propose a transparent model distillation approach to detect bias" in black-box risk scoring models.</p>
<p>Paul Voosen, <i>How AI Detectives are Cracking Open the Black Box of Deep Learning</i>, Science (July 6, 2017, 2:00 PM), http://www.sciencemag.org/news/2017/07/how-ai-detectives-are-cracking-open-black-box-deep-learning.</p>	<p>Author examines different ways researchers/scholars are tackling the interpretability problem of AI to understand how neural networks make decisions.</p> <p>"That interpretability problem, as it's known, is galvanizing a new generation of researchers in both industry and academia. Just as the microscope revealed the cell, these researchers are crafting tools that will allow insight into the how neural networks make decisions. Some tools probe the AI without penetrating it; some are alternative algorithms that can compete with neural nets, but with more transparency; and some use still more deep learning to get inside the black box."</p>

GREG WALDEN, OREGON
CHAIRMAN

FRANK PALLONE, JR., NEW JERSEY
RANKING MEMBER

ONE HUNDRED FIFTEENTH CONGRESS
Congress of the United States
House of Representatives
COMMITTEE ON ENERGY AND COMMERCE
2125 RAYBURN HOUSE OFFICE BUILDING
WASHINGTON, DC 20515-6115
Majority (202) 225-2927
Minority (202) 225-3641
December 15, 2017

Dr. Michael Kearns
Computer and Information Science
University of Pennsylvania
509 Levine Hall
3330 Walnut Street
Philadelphia, PA 19104

Dear Dr. Kearns:

Thank you for appearing before the Subcommittee on Communications and Technology and the Subcommittee on Digital Commerce and Consumer Protection on Wednesday, November 29, 2017, to testify at the joint hearing entitled "Algorithms: How Companies' Decisions About Data and Content Impact Consumers."

Pursuant to the Rules of the Committee on Energy and Commerce, the hearing record remains open for ten business days to permit Members to submit additional questions for the record, which are attached. The format of your responses to these questions should be as follows: (1) the name of the Member whose question you are addressing, (2) the complete text of the question you are addressing in bold, and (3) your answer to that question in plain text.

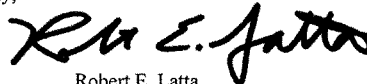
To facilitate the printing of the hearing record, please respond to these questions with a transmittal letter by the close of business on Tuesday, January 9, 2018. Your responses should be mailed to Evan Viau, Legislative Clerk, Committee on Energy and Commerce, 2125 Rayburn House Office Building, Washington, DC 20515 and e-mailed to Evan.Viau@mail.house.gov.

Thank you again for your time and effort preparing and delivering testimony before the Subcommittees.

Sincerely,



Marsha Blackburn
Chairman
Subcommittee on Communications
and Technology



Robert E. Latta
Chairman
Subcommittee on Digital Commerce
and Consumer Protection

cc: The Honorable Michael F. Doyle, Ranking Member, Subcommittee on Communications and Technology

The Honorable Janice D. Schakowsky, Ranking Member, Subcommittee on Digital Commerce and Consumer Protection

Attachment

**RESPONSES OF DR. MICHAEL KEARNS
QUESTIONS FOR THE RECORD
HEARING ON “ALGORITHMS: HOW COMPANIES’ DECISIONS ABOUT DATA
AND CONTENT IMPACT CONSUMERS.”
SUBCOMMITTEE ON COMMUNICATIONS AND TECHNOLOGY AND THE
SUBCOMMITTEE ON DIGITAL COMMERCE AND CONSUMER PROTECTION
HOUSE COMMITTEE ON ENERGY AND COMMERCE**

The Honorable Robert E. Latta

1. How is information that companies obtain about consumers used to model an individual and make predictions about a person’s behavior?

A. Detailed information about consumers (gender, age, race, location, app usage, search queries, etc.) is analyzed using modern machine learning and statistical methods to create highly individualized predictive models. These models may discover informative combinations of the inputs that would be difficult for humans to discern, and may make predictions about consumers that are much more revealing than the raw data. Large and diverse data sets about consumers are the foundation for effective models that predict behavior.

a. What are the benefits and dangers of these models?

A. These models present a number of substantial benefits to both consumers and advertisers. For example, these models help consumers make purchasing decisions based upon their interests, and they help marketers target advertisements to consumers who are likely to be interested in their products. The potential dangers of these models involve their impact on (1) consumer privacy, (2) the incentive that online companies have to amass massive amounts of information about consumers, and (3) the potential for bias or discriminatory behavior based upon the information produced by the models.

2. Can real or perceived bias be cured in highly complex algorithmic systems to enhance reliability or intended outcomes?

a. If so, how do online platforms conduct this curing or correction process?

A. In the past few years, new algorithmic research has emerged that may provide practical methods to reduce bias, while still achieving good predictive accuracy. These methods include algorithms for auditing predictive models for bias, and reducing or correcting that bias. Online platforms are still in the process of considering implementing such methods into their modeling.

3. There are conflicting reports about how accurately companies can predict what consumers are really interested in. For instance, the Wall Street Journal wrote in November 2017 a story entitled “Google Has Picked An Answer For You—Too Bad It’s Often Wrong.” In that article, they note “the Internet giant is promoting a

single result over all others, and many are contentious, improbable or laughably incorrect.” Are we still in the early stages of companies being able to accurately gauge their users’ interests?

A. That particular article is talking about something quite different than modeling consumer behavior, preferences and desires. Instead, it is focused on more of a pure language understanding problem, such as answering questions like “Does money buy happiness?” or “Who are the worst CEOs of all time?” For these questions, the massive amounts of data that companies like Google, Amazon, and Facebook have collected on consumer behavior, at the collective and individual level, is not especially helpful. So, for example, your Amazon purchases, GPS coordinates and Google searches are incredibly valuable in predicting your future online and offline behavior, but they do not help answer the question about money buying happiness.

a. Are there any special considerations for Internet service providers?

A. As I mentioned in my testimony, in general, the large consumer-facing tech companies have amassed large and diverse data sets that are directly relevant to making detailed inferences about individuals, including search queries, shopping behavior, location data, and social interactions. ISPs generally do not have access to the same depth or breadth of data, in part because of packet encryption via the https protocol.

The Honorable Gregg Harper

1. Based on your research, do you think consumers do things they otherwise would not because of ho[w] their data is being used? Or are they instead being presented options that they may not have known they had?

A. Consumers are definitely presented with options they would not have had if their data was not collected and analyzed by online platforms. In general, every Amazon recommendation you receive, and the ads you see on Google, are tailored and specialized based on your particular past behavior. Sometimes, this specialization may present users with beneficial choices — as in when Amazon recommends a book I would love that I didn’t even know about. And sometimes these choices may be detrimental or even discriminatory, as when ads for high-interest payday loans are targeted towards low-income individuals.

The Honorable Michael C. Burgess

1. If a digital platform or intermediary knows all of a user’s travels over the course of weeks or months, is there anything they can’t deduce by correlating location with mapping? For example, can illness be inferred if a person is repeatedly going to an out-patient facility?

A. In general, GPS and other precise location-based data is tremendously powerful, especially when combined with other public or commercial data sets mapping “points of interest” (stores, parks, medical facilities, homes, etc.) to physical location. The example you give is more than plausible — location data showing a consumer repeatedly visiting a

chemotherapy facility might indeed strongly suggest that person, or someone close to them, is receiving treatment.

The Honorable Adam Kinzinger

1. Could a malicious actor hack and combine the collected data to create a full profile of an individual and use that profile to access their accounts?

A. Protecting the security of consumer information must be a priority. A malicious actor could obtain detailed profiles of people by hacking into a data source or company with sufficiently rich and diverse consumer information. It's conceivable that a hacker could develop a level of knowledge about a consumer to answer basic account security questions or correctly determine a consumer's password.

2. How good of a model of an individual can be made from existing data sources?

A. The largest consumer-facing tech companies have amassed massive data sets that have enabled them to develop very accurate models of individual consumers.

a. Do you think it is possible to predict the wants of a consumer before that consumer knows them?

A. Yes, machine learning and algorithms are widely used by consumer-facing technology companies to predict consumer purchasing habits, potential social connections, and search queries.

2. How do algorithms reflect the biases and interests of their creators? Is it possible to write non-normative code?

A. In general, I do not believe there is a widespread problem of programmers encoding their personal beliefs and biases directly into their algorithms. But there is no shortage of other ways in which deployed algorithms and models demonstrably exhibit biases (by age, gender, race, etc.). Regarding writing non-normative code, there is indeed a recent but growing body of research providing design principles for making algorithms less biased.