Data Justice:
Taking on Big Data as an Economic Justice Issue

Nathan Newman

March 2015
Nathan Newman is Director of Data Justice and has been writing for twenty years about the impact of technology on society, including his 2002 book Net Loss: Internet Profits, Private Profits and the Costs to Community, based on doctoral research about rising regional economic inequality in Silicon Valley and the nation and the role of Internet policy in shaping economic opportunity. This report builds on his research about big data as a research fellow at the New York University Information Law Institute from 2012 to 2014, including authoring two law reviews this year on the subject. Those papers include "The Costs of Lost Privacy: Consumer Harm and Rising Economic Inequality in the Age of Google" and published in the William Mitchell Law Review, highlighted the economic harm to consumers and rising economic inequality stemming from big data. The second, entitled "Search, Antitrust and the Economics of the Control of User Data" and published in the Yale Law Journal on Regulation, detailed the case for antitrust action against Google based on its control of user data and harm to consumers. He has a J.D. from Yale Law School and a Ph.D. from UC-Berkeley’s Sociology Department.

Data Justice is a newly launched project to promote public education and new alliances to challenge the danger of big data to workers, consumers and the public. Data Justice will educate key stakeholders, allies, and the public on approaches to prevent big data platforms from using that data in ways that harm consumers. We will highlight the way big data platforms facilitate the exploitation of employees, consumers and citizens by abusive financial services companies and thereby increase economic discrimination and inequality in the economy.

http://www.datajustice.org
@datajustice1
# Table of Contents

I. Introduction: Why Big Data is an Economic Justice Issue, Not Just a Privacy Issue, for Individuals and Society .................................................................1

II. How Big Data Platforms Harm Workers, Consumers and Society ..........................................................3

   How Big Data Harms Workers and Undermine Wages ..........................................................3
   Big Data Platforms Buy and Sells User Data — and Drive Consumer Profiling ..................4
   What Advertisers Get for Their Money: Knowing What People Will Buy and the Price At Which They Will Buy It ........................................................................5
   Big Data Platform Price Discrimination Increases Prices Overall for Consumers ..........6
   Big Data Platforms Enable Racial Profiling and the Exploitation of the Most Economically Vulnerable Groups in Society ..........................................................6
   Big Data Platforms’ Helped Facilitate the Subprime Mortgage Debacle and Its Aftermath ...8
   Consumers Lose Financially as the Value of their Personal Information Flows to Big Data Platforms ..................................................................................................9
   Consumers Lose Access to Local Journalism as Advertising is Diverted to Big Data Companies .................................................................................................11

III. How Big Data Reinforces Market Power .................................................................13

   Network Effects Lock-in Market Power by Big Data Platforms ..................................................13
   Data Platforms’ Control of User Information Creates Barriers to Entry for Potential Competitors Lacking That Data .................................................................14
   Controlling More Data Increasingly Trumps Other Forms of Innovation .........................14
   Control of Unique Data Strengthens Control by Big Data Platforms .....................................15
   Why Control of User Data Matters: Users are the Product, Not the Consumers, of Advertising-Supported Data Platforms .........................................................16
   Expanding into Multiple Related Sectors Reinforces Market Power in Core Data Platform Businesses .........................................................................................17
   Google as a Focus for Initial Action Among Big Data Platforms .........................................18
   Lack of High-Tech Competition Highlights Need for Action by Regulators ....................19

IV. Taking Action to Protect Workers, Consumers and Society in an Era of Big Data Platforms .................................................................21

   Promote “Self-Help” Campaigns to Empower Consumers to Take Control of their Own Data .................................................................................................21
   Limit Employer Profiling in Hiring and of Current Employees ............................................21
   Strengthen Regulations and Laws Protecting Consumer Privacy ..................................22
   Restructure Markets to Promote Competition ..................................................................23
   Directly Regulate and Restrict Practices that Harm Consumers ......................................23

V. Conclusion: Stemming Rising Economic Inequality by Better Regulation of Big Data Platforms .................................................................25
Executive Summary

I. Why Big Data is an Economic Justice Issue, Not Just a Privacy Problem

A. “Big Data” platforms such as Google and Facebook are becoming dominant institutions organizing information not just about the world but about consumers themselves, thereby reshaping a range of markets based on empowering a narrow set of corporate advertisers and others to prey on consumers based on behavioral profiling.

B. This control of personal data is not just an issue of privacy but of economic justice; if information is the coin of the new online economy, this steady loss of data by individuals is in turn driving broader economic inequality in a host of ways.

C. There is a particular concern that “free services” on the Internet use consumer data for the benefit not of those users but for third party corporate customers of those data platforms, particularly advertisers who drive a large portion of the revenue model of the online Internet economy.

D. Control of more data seems to be the key to dominating particular economic sectors and such dominance in turn yields more data from consumers. The dynamics of big data is fueling market concentration in industry sectors across the economy and consumers lose out as dominant institutions have even less pressure to share the economic value of personal data with those users.

II. How Big Data Platforms Harm Workers, Consumers and Society

A. While big data can benefit workers and consumers in certain instances, there are a range of new consumer harms to users from its unregulated use by increasingly centralized data platforms.

B. These harms start with the individual surveillance of users by employers, financial institutions, the government and other players that these platforms allow. The most direct fear many people have is that employers, schools, insurance companies and others will access information on data platforms in ways that deny them jobs or services.

C. More deeply, new “algorithmic profiling” of groups of people allow corporate institutions to discriminate and exploit consumers as categorical groups.

D. Behavioral profiling allows advertisers to offer goods at different prices, what economists call price discrimination, to extract the maximum price from each individual consumer. Such online price discrimination raises prices overall for consumers, while often hurting lower-income and less technologically savvy households.

E. Behavioral profiling is used by especially seedy companies to target a variety of financial and economic scams at vulnerable populations most likely to fall prey to their offers. Examples include subprime mortgages targeting vulnerable consumers with worse deals based on racial and economic profiling. Advertising-driven platforms continued to benefit from scam “mortgage modification” and payday lender advertisers exploiting financially distressed households in the wake of the financial crisis.

F. User data is economically valuable, yet big data platforms manage to extract data from users with little financial compensation. Data platforms’ revenue is based first on harvesting the largely free labor providing user-provided content on the web (search, videos, reviews on Amazon, shared social
III. How Big Data Reinforces Market Power

A. As a handful of data platforms generate massive amounts of user data, the barriers to entry rise since potential competitors have little data themselves to entice advertisers compared to the incumbents who have both the concentrated processing power and supply of user data to dominate particular sectors. Increasingly, analyzing the maximum quantity of data and finding useful correlations within that data is trumping other forms of insight and innovation, making it even harder for any startup with the proverbial better idea to win out against the pure quantities of data held by incumbents.

B. As companies collect user data, they gain competitive advantage against any potential challenger who will lack access to that unique user data in setting up any rival service. Such data can be redeployed by dominant players not just to strengthen their position in existing services but used in related new services to expand their economic reach. As data platforms expand into related online sectors and acquire promising startups, the big data platforms gain access to additional unique data that leaves potential competitors with even less openings to challenge their market dominance.

C. When thinking about competition in these advertising-driven data markets, the point is not how easy it is for users to switch to an alternative product but whether advertisers can easily substitute an alternative online advertising service for a dominant one. Users are the product, not the consumers, of advertising-supported data platforms. One example of this market power from controlling user data is the premium price Google is able to demand from advertisers.

D. Aside from the broader harm to society and consumers outlined early in this report, the most basic consumer harm that should demand a regulatory response are the inevitable higher costs to the advertising customers, which in turn are inevitably passed onto to consumers. What makes the harm to average consumers from current dominance by big data platforms seem distinctly worse is that the higher profit margins for those data platforms do not seem to be plowed back into providing equally valuable services for end users in the same way advertising in traditional media was reinvested into media content.

E. By spreading into multiple related data-driven business sectors, the dominant data platforms reinforce market power in their core businesses. Expansion into new sectors adds to their overall storehouse of user data, while choking off potential sources of data for potential competitors.

F. While the danger to consumers and the prospect of increasing economic inequality discussed in this paper is reflected in dynamics across the online (and increasingly offline) economy, Google could be a key initial target for advocates and regulators to focus on since it is in many ways the leading edge of a broader set of data platforms.

G. Contra arguments that antitrust and other pro-competition regulation is outmoded in the new, high-tech economy, the reality is that the rise of the Internet has coincided with a drastic decline in new competitors emerging to challenge dominant firms. With the market is producing fewer likely
challengers to keep dominant data platforms in check, it becomes even more important for regulators and courts to take action earlier in technology markets, before market dominance arises.

IV. Taking Action to Protect Consumers in an Era of Big Data Platforms

A. A first step for advocates may be to strengthen and expand on ongoing campaigns that seek to rein in data platforms through boycotts and promoting “self-help” technological tools to protect consumer privacy. However, encouraging “self-help” actions by consumers is not likely to make a significant difference in actually strengthening consumer power over their data, since the determined tracking of users by advertising largely frustrates such tools.

B. Mandating greater privacy, data portability, and transparency is a strong first step in protecting consumers, but needs a real regime of regulatory and court enforcement against companies violating such anti-tracking and anonymity rules protecting consumers. Requiring explicit consent for all uses of user data could jumpstart a real market for user data. This could open up more space for new companies to compete on incentives at that point of friction and potentially encourage all data platforms to either better protect privacy or share more of the profits of the industry directly with users.

C. Federal regulators should be looking at how to structure data platform sectors to both promote more competition and encourage more consumer power within those sectors. State attorneys general could also be approached (with the goal of enlisting the US FTC or Justice Department in the long term) to support an antitrust lawsuit against Google around its overall dominance of the search-advertising sector.

D. However, encouraging more competition for data platforms and empowering individual users to control their data will still fail to protect many consumers, especially in areas of consumer harm where the danger is indirect, in the future, or deliberately concealed. One clear step would be to bar data platforms from engaging in price discrimination or from knowingly facilitating price discrimination where different groups are secretly offered different prices by its advertisers for the exact same product or service. Regulators could also bring the participation of big data platforms in marketing financial services under the regulation of the Consumer Financial Protection Bureau.

IV. Stemming Rising Economic Inequality by Better Regulation of Big Data Platforms

A. As more of the economy moves online, the importance of data mining and the asymmetry of control of information becomes ever more critical in economic markets. Information asymmetry between big data companies and consumers is easily converted into economic inequality when one side of every transaction has so much more knowledge about the other during bargaining.

B. The increasing information asymmetry in consumer markets, driven by data mining and facilitated by online services, may be an additional significant cause of this overall increase in economic inequality we have seen over the last four decades.

C. Government authorities using regulatory tools can stem at least part of this trend by restoring a degree of control by individuals over what personal data is shared online and the financial terms on which that data is shared. Such pressure might translate into a greater focus on sharing the financial bounty of user information with those users, serving both equity and competition.

D. If nothing else, public advocacy for these changes can be a chance for a much broader public debate on the abuses of data mining online and how to make all markets work more fairly for average working families.
I. Introduction: Why Big Data is an Economic Justice Issue, Not Just a Privacy Issue, for Individuals and Society

Data has been called the “new oil” of the information age, an asset used by corporations to reshape markets and increase their market power and profits.¹ On the Internet, we see the rise of new “big data” platforms such as Google, Amazon, Apple, Facebook and others that accumulate ever increasing information on consumer behavior, interests and needs. While this data unquestionably increases the efficiency of the economy in numerous ways, what is in question is whether consumers are ultimately benefitting significantly from those productivity gains or whether that surplus is being largely captured by these big data platforms themselves. Worse, the increasing loss of control of private data by individuals seems to be leaving individuals vulnerable to economic exploitation by a range of corporate actors. 

Big data is often framed as the promise of economic gain for society versus the problem of loss of privacy, but the reality is that it is precisely how big data may be making many people less well off that should be the greatest cause for concern. Unless the power of big data platforms is curbed in favor of more individual and democratic control of the use of personal data, then big data needs to be considered as much an issue of economic justice as a privacy issue.

These big data platforms, what Jaron Lanier calls “siren servers” in his book, Who Will Own the Future?,² attract consumers with a variety of services that encourage those users to part with personal data, which in turn is analyzed and combined with private information from other users in massive networks of computers. These companies use that analysis to reshape markets – “disrupt” in Silicon Valley parlance – and channel an ever greater share of economic wealth into the hands of these big data platforms.

Characterizing the collection of personal data as a “loss of privacy” for consumers therefore drastically underplays what is at stake in the growing control of data by these dominant corporate players. Fundamentally, if information is the coin of the new online economy, this steady loss of data by individuals is in turn driving broader economic inequality in a host of ways.

There is a particular concern that “free services” on the Internet use consumer data for the benefit not of those users but for third party corporate customers of those data platforms, particularly advertisers who drive a large portion of the revenue model of the online Internet economy. While much of that advertising no doubt serves traditional advertising goals of strengthening brand awareness or promoting new products to consumers, the rise of behavioral profiling of consumers using the private data extracted by these big data platforms increases the use of advertising for more exploitative practices.

Big data platforms facilitate advertisers engaging in user profiling that aids those companies in extracting the maximum profit possible from consumers in the overall economy. Advertisers can deliver ads not just to the users most likely to be interested in the product but can tailor prices for individual consumers in ways that can maximize the revenue extracted from each purchaser. Consumers can be profiled and offered higher prices, unaware that other customers are getting better deals, while financially struggling households are tagged as vulnerable and offered economically exploitative services such as payday and subprime loans.

It is largely because of the ability to profile users and more precisely target ads that online advertising as a whole has exploded and become
the largest advertising sector in the United States. In fact, 2013 was the year Internet advertising surpassed broadcast advertising revenues in the United States for the first time.\(^3\) Online advertising amounted to $42.8 billion in the United States\(^4\) and $117.2 billion globally\(^5\) that year.

Since the rise of big data has coincided with the stagnation of incomes for average households, policy makers should be raising concerns that, alongside traditional explanations of rising inequality such as deunionization, globalization, and automation of unskilled jobs, the concentration of data into ever fewer corporate hands is helping to drive economic inequality in the broader economy.

While big data can benefit consumers in certain instances, regulators need to take action to address new consumer harms to users from its unregulated use by increasingly centralized data platforms. The Federal Trade Commission has itself highlighted some of these problems in a number of recent reports,\(^6\) as well as litigation against companies engaged in deception in collecting personal data,\(^7\) but it is clear that additional regulation and laws are needed to address the full scope of the harm to consumers.

These harms start with the individual surveillance of users by employers, financial institutions, the government and other players that these platforms allow, including denial of employment or scholarships based on what people post to their personal social media sites. While a few states have taken action to restrict misuse of social media data to directly, there is a broader need for federal action.

However, such individual surveillance may be less of a danger to consumers than the broader aggregation of data – so-called “algorithmic profiling” -- and the ways it allows companies to discriminate and exploit consumers as categorical groups. Much of this profiling is invisible to consumers, making the need for public action all the more urgent and needed. Big data platforms collect so much information about so many people that correlations emerge that allow users to be slotted into marketing categories in unexpected and often unwelcome ways. Increasingly, every transaction, every website viewed, and every action online generates a data trail swept into the data platforms online. Most websites invite dozens of companies to track users on their site and follow them across the web.

Control of more data seems to be the key to dominating particular economic sectors and such dominance in turn yields more data from consumers, so the dynamics of big data is fueling market concentration in industry sectors across the economy. Not only does control of user data in a sector create barriers to entry against new firms, companies with large troves of unique data about consumers more easily repurpose that data for new uses that allow them to dominate new industry sectors.

Since the rise of big data has coincided with the stagnation of incomes for average households, social justice advocates should be raising concerns that, alongside traditional explanations of rising inequality such as deunionization, globalization, and automation of unskilled jobs, the concentration of data into ever fewer corporate hands is helping to drive economic inequality in the broader economy. Given the direct consumer harms and economic inequality being driven by use of that data, the argument of this report is that taking on the power of these big data platforms should be a far higher priority for economic justice advocates and those concerned about the problem of rising economic inequality.
II. How Big Data Platforms Harm Workers, Consumers and Society

While big data can benefit consumers in certain instances, there are a range of new consumer harms to users from its unregulated use by increasingly centralized data platforms. These harms start with the individual surveillance of users by employers, financial institutions, the government and other players that these platforms allow, but more deeply extend to forms of what has been called "algorithmic profiling" that allow corporate institutions to discriminate and exploit consumers as categorical groups.

How Big Data Harms Workers and Undermine Wages

The most direct fear many people have is that employers, schools, insurance companies and others will access information on data platforms in ways that deny them jobs or services. Horror stories already abound:

- A teacher was fired for complaining about getting sick from illnesses caught from her students after jokingly referring to her kids as “germ bags” on Facebook.  
- A judge allowed Wal-Mart to subpoena materials from Facebook, Myspace and Meetu.com to challenge an employee claim of work-related injury causing him head and neck pain.

In fact, a survey found that one in ten young people have been rejected for a job because of their social media profile. Many schools report using Google, Facebook, YouTube and other sites to check out student sin determining scholarships. For many companies, their job monitoring of potential and current employees is assisted by companies like Spokeo, a search technology that compiles information about millions of identifiable individuals from other social networks and hundreds of other online and offline sources, including real estate listings and marketing surveys. The company has featured banners inviting “HR Recruiters – Click Here Now!” to get information about potential hires.

Credit checks and other scouring of online background information are thereby being combined into algorithmic screening categories that may exclude individuals from even being considered for a job.

Employees also increasingly face prospective and current employers demanding that they share passwords on personal sites such as their social networking accounts which erode and even breaks basic divisions between work and personal life in ways that obliterate the privacy of employees. By revealing a religious, ethnic and other information about candidates and employees, this exposure and use of social media data increases the likelihood of other discriminatory action against an employee.

Employers have also turned to pre-hiring personality tests as a method to engage in a range of discriminatory practices and to avoid hiring people who might challenge illegal employment practices in the workplace, the latter a form of pre-hire retaliation against workers who might act to protect their rights. While courts have sometimes found that use of many personality tests are illegally used to identify applicants subject to depression or other mental health diseases in violation of the Americans with Disability Act or have also found such tests intrude into personal issues of religious belief and sexual orientation, they are still used in other unfair and biased ways. Such tests often do little
to make valid assessments of the particular skills an applicant may have but instead are often used merely to identify individuals who will not challenge abusive or even illegal treatment in the workplace. Such tests also intrude into the personal lives and privacy of individuals in ways irrelevant to their job performance and should be restricted.15

Big Data Platforms Buy and Sells User Data —and Drive Consumer Profiling

More broadly, big data platforms collect so much information about so many people that correlations emerge that allow users to be slotted into marketing categories in unexpected and often unwelcome ways. In one notorious example, the New York Times revealed that companies like Target had found that pregnant women purchased specific combinations of unscented lotion and certain supplements, allowing the creation of a “pregnancy prediction” score to allow marketing of baby-related goods to women who had not publicly revealed to anyone that they were expecting.16 A data broker like Acxiom assigns a 13-digit code to each person and assigns them to 70 “clusters” based on traits such as age, income, race, whether you live in an urban or suburban community, and where you shop.17

Increasingly, every transaction, every website viewed, and every action online generates a data trail swept into the data platforms. Most websites invite dozens of companies to track users on their site and follow them across the web. One study found that in 2013, there were 328 separate companies tracking visitors on the top fifty content Web sites.18 Internet providers like Comcast automatically install tracking tools on their own customers. Each site a consumer uses usually deposits a tiny bit of code on their hard drive, called a “cookie” which allows companies to track and aggregate the overall activity of consumers across the whole Internet. The placement of such online cookies on tens of thousands of website was pioneered by online advertiser DoubleClick, which was acquired by Google in 2007, and allows online advertisers to display different banner ads on any site based on the profile generated by this aggregate data.19 Companies often bid furiously for the right to place their banner for a particular user. And the ads can follow you around the web; for example, if you view a camera on eBay, companies will bid in an auction for your “cookie” and then display ads related to digital cameras on other web pages you visit.20

With the rise of smartphones, advertisers could increasingly sell mobile ads based on the location of any consumer. Apple, Google and mobile phone operators like Verizon and AT&T control data that is of extreme interest to advertisers,21 since where people live, work and shop often says more about who you are than any other single piece of data. Increasingly, off-line behavior beyond a person’s location is also being combined with online activity to maximize the profiling of users. For example, Facebook has combined demographic and shopping information from data brokers like Acxiom with its own social media-generated data as well as corporate advertisers’ consumer data to better target ads on its network.22 If consumers buy a product using a loyalty card at a particular store, an ad reflecting that purchase is now likely to show up in their Facebook feed.23

The dynamics of online display advertising described above is actually only a minority (19%) of the overall online advertising sector, where Google and Facebook dominate the sector but share revenue with a larger number of players. Far larger is the search-advertising sector (41% of all online advertising) where Google controls most of the sector and derives an even higher percentage of the revenue. In search advertising, the core product sold are the little text advertisements that appear next to search results, on Gmail pages, and on a range of other Google and affiliated sites across the internet. Particular ads are tied to particular words users type into search engines or based on words on a particular webpage where the advertisements appear. This is combined with whatever other data the search engine has about the particular viewer, from emails sent to friends and colleagues, videos previously viewed, previous online searches, their location and any other interests, demographic information or other behavioral characteristics that the advertising platform knows about the user.24

Advertisers bid in an auction on these keywords for each any of the desired behavioral and demographic sub-groups that the advertiser is trying to target. A combination of the highest bid price and the quality of the advertiser’s site.25
determines which ad gets the most prominent placement on the search or affiliated page. An innovation in search advertising is that advertisers don’t pay a dime unless a viewer actually clicks on the ad.

It is largely because of the ability to profile users and more precisely target ads that online advertising as a whole has exploded and become the largest advertising sector in the United States. In fact, 2013 was the year Internet advertising surpassed broadcast advertising revenues in the United States for the first time. Online advertising amounted to $42.8 billion in the United States and $117.2 billion globally that year.

**What Advertisers Get for Their Money: Knowing What People Will Buy and the Price At Which They Will Buy It**

The question is what do advertisers get for their money? No doubt, user profiling helps advertisers more effectively identify the customers most likely to be interested in their products. However, the darker explanation is that such profiling also facilitates tailoring prices to individual consumers in ways that maximize revenue extracted from each transaction.

This ability to charge different prices to different customers for the same good or service, what economists call “price discrimination”, is based on the reality that people have different maximum prices they are willing to pay. And profiling consumers helps advertisers identify this “pain point” for each customer and offer a different price to each customer matching that maximum price they are willing to pay without them knowing that other deals are available. Some economists argue that where consumers know all pricing options, they can potentially benefit from price discrimination, as when airline passengers choose between a cheap price at an inconvenient time to save money, which can fill seats, increase revenues for airlines and increase options for different customers. But when people either don’t know about better deals or don’t easily have the ability to access them, such price discrimination is far more likely to hurt consumers.

For example, a 2012 *Wall Street Journal* report found that major companies, including Staples, Home Depot, Discover Financial Services and Rosetta Stone, were systematically using information on user physical locations to display different online prices to different customers. More disturbingly, contrary to any hope this might benefit low-income bargain hunters, the paper found that higher-income locations were offered better deals than low-income communities, because those poorer areas had fewer local retail outlets competing with the online stores. Credit card companies like Capitol One show different offers with different credit card deals based on view locations and guesses by the company about their income.

In search advertising, this differential pricing overwhelmingly takes the form of web coupons offered to some people but not others based on their behavior and demographic data. As Ed Mierzwinski, consumer program director for the United States Public Interest Research Group (USPIRG) noted in an interview, companies “offer you, perhaps, less desirable products than they offer me, or offer you the same product as they offer me but at a higher price.”

Economists like Nobel Prize Winner Joseph Stiglitz, who pioneered what has been called “information economics”, detail the economic harm to consumers from such differential pricing. When consumers don’t know all their price options, it creates “market power in product markets” which firms exploit through sales and “other ways of differentiating among individuals who have different search costs” in identifying different price options.
Big Data Platform Price Discrimination Increases Prices Overall for Consumers

Many people had a vision of an online economy where consumers could quickly compare prices, but studies have shown that hidden discounts, the posting of multiple versions of the same product, and other “price obfuscation” strategies are designed to frustrate consumers and keep prices up. Where prices are obscured and sellers impose price discrimination, economic models generally show that overall prices in the economy will end up higher than any model where consumers knew all prices.

This argument is not one initially made by critics of the online economy but has actually been made by boosters of the opportunity for companies to profit from it. Academic Hal Varian has a long history of examining various models of price discrimination and in 2005, he was appointed Chief Economist for Google. That same year, he co-authored an article in the industry-based academic journal Marketing Science touting the gains for companies engaging in online price discrimination, particularly against what the authors labeled “myopic” consumers who are unaware of how their data is used to structure different prices for them.

Varian and his coauthor argued that “significant initial investments in information technology can lead to competitive advantages” that lock-in user loyalty while collecting personal information to make price discrimination profitable. In a foreshadowing of both Google’s and other data platforms’ practices, the article argued for companies to lock-in users to particular services, block anonymous participation, and seek out the coveted “myopic” consumers to increase profits.

While various economic models yielded different results in Varian’s and his co-author’s analysis, they generally agreed that, in many cases, any economic value added to the economy due to increased efficiencies “is entirely due to the increased profit received by the seller” while in other cases, consumer welfare actually falls overall. In particular, the “myopic” consumers generally lose out financially under price discrimination using targeted consumer profiling.

Recent research on online advertising reinforces this analysis of consumer loss due price discrimination combined with consumer profiling. Comparing traditional regimes of mass-market advertising to online advertising, researchers Rosa-Branc Esteves and Joana Resende found that average prices with mass advertising were lower than with targeted online advertising. Similarly, Benjamin Reed Shiller found that where advertisers know consumers willingness to pay different prices, companies can use price discrimination to increase profits and raise prices overall, with many consumers paying twice as much as others for the same product.

Big Data Platforms Enable Racial Profiling and the Exploitation of the Most Economically Vulnerable Groups in Society

Once upon a time, people celebrated the Internet as promising a new era where shoppers, invisible on the web, could not be judged based on their race or otherwise discriminated against. However, online behavioral targeting can combine a home address and a few more characteristics to create an almost perfect proxy for race. If anything, such online discrimination can be more vicious for its subtlety and invisibility since customers don’t even know what prices are being offered to other people of different races or socioeconomic circumstances. And it’s not even clear that current laws could fully address such harms if they could be made visible, since as George Mason University professor Rebecca Goldin noted in a 2009 article, what would be the legal status if banks used “the kind of music one buys to determine his or her loan rate?”

Such online “weblining” has been well documented. Along with the price discrimination based on location discussed above, companies like Wells Fargo listing houses for sale have collected...
zip codes of online browsers and directed those buyers towards neighborhoods of similar racial makeup. This online discrimination parallels the broader reality of companies like Wells Fargo illegally steering an estimated 30,000 black and Hispanic borrowers from 2004 to 2009 into more costly subprime mortgages or charging them higher fees than comparable white borrowers.

As ColorLines magazine has noted, a "user’s browsing history, their location and IP information...combined with information available in Google’s public data explorer (including US census, education, population, STD stats, and state financial data) presumably could also be folded into the personalized search algorithm to surmise a lot more than your race." Latanya Sweeney described in an academic article how, on sites detailing legal information about individuals, when people searched for a name "on the more ad trafficked website, a black-identifying name was 25% more likely to get an ad suggestive of an arrest record." What is disturbing is that people online can find themselves losing opportunity as their ongoing behavior or interests lump them in with the "wrong" racial or other group in the algorithms of big data platforms. For example, Kevin Johnson, a condo owner and businessman, found that after returning from his honeymoon, his credit limit had been lowered from $10,800 to $3800. The change was not based on anything he had done but, according to a letter from the credit card company, he had shopped at stores whose patrons “have a poor repayment history.” If your habits associate you with particular categories or groups, you will invisibly find opportunities opening up or closing down based on how data algorithms choose to place you. Similarly, whether you get a refund when making a complaint to a company will often be heavily influenced by the categories in which data analysis places a caller.

For less ethical companies, big data gives them the ability to seek out the most vulnerable prospects to exploit and entice them with scams and misleading offers. Such niche scams and economically exploitive relationships can be focused on those most vulnerable to the scam’s

**Table I: Company Product Names**

<table>
<thead>
<tr>
<th>Sample List of Targeting Products Identifying Financially Vulnerable Populations</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Burdened by Debt: Singles&quot;</td>
</tr>
<tr>
<td>&quot;Mid-Life Strugglers: Families&quot;</td>
</tr>
<tr>
<td>&quot;Resilient Renters&quot;</td>
</tr>
<tr>
<td>&quot;Very Spartan&quot;</td>
</tr>
<tr>
<td>&quot;X-tra Needy&quot;</td>
</tr>
<tr>
<td>&quot;Zero Mobility&quot;</td>
</tr>
</tbody>
</table>

Source: Company Responses
appeal, while remaining essentially invisible to everyone else, including reporters and researchers trying to evaluate the harms from online advertising methods.

The data broker industry even has a term – “sucker lists” – for the poor, old and less educated groups that they compile for such unethical marketers. For example, people who reply to sweepstakes offers are put onto a list by one data broker and offered to advertisers as an “ideal audience for... subprime credit offers” and other enticements. Other lists include “suffering seniors” who are identified as having Alzheimer’s or similar maladies.48 (See the list on the previous page from a chart in the Rockefeller Senate Data Broker report) The Federal Trade Commission itself has noted that when companies use a consumer’s financial status to send targeted advertisements, it is not covered by FCRA if they don’t cover specific pre-approved offers of credit.49

Search advertising is especially attractive to companies looking for micro markets of vulnerable targets for scams, since the combination of keyword searches and demographic data allows what writer Jaron Lanier calls the “ambulance chasers and snake oil salesmen” of the Internet to get targeted access to victims. The “minimalist link” of a search ad focuses on lead generation for such companies where users self-select into the advertisers’ target group by clicking on the link.50 For example, one company ran advertisements for poisons and chemicals on the Google Group page alt.suicide.methods where users were discussing how to kill themselves.

Reflecting the more comprehensive problems in search advertising targeting the vulnerable, Google in August 2011 agreed to pay a $500 million civil forfeiture to the federal government, one of the largest in history, as part of a settlement for the company knowingly allowing illegal pharmacies to target users on its search engine.51 The company had been put on notice by the government as early as 2003 that companies were selling illegal steroids and fake prescription medicine to desperately ill individuals, yet the company not only accepted the ads but its staff helped foreign-based pharmacies write their ads for maximum effectiveness. It was only when a felon, David Whitaker, collaborated with the government in a sting operation that the full extent of the company’s collaboration with such scam and illegal marketers was fully documented, including that knowledge of the collaboration went all the way up to CEO Larry Page.52

**Big Data Platforms’ Helped Facilitate the Subprime Mortgage Debacle and Its Aftermath**

Big data lay at the heart of the subprime mortgage and overall financial meltdown the nation suffered at the end of the last decade. Data crunchers were key to manipulating financial markets and securities throughout the financial industry and big data platforms were critical parts of the marketing machine that pushed subprime financial products out to the most vulnerable members of the American public.

In fact, by the mid-2000s, the lion’s share of the online advertising economy was being driven by subprime and related mortgage lenders. As Jeff Chester of the Center for Digital Democracy said back in 2007 “Many online companies depend for a disproportionate amount of their income on financial services advertising, with subprime in some cases accounting for a large part of it.”53 As the subprime frenzy was hitting its height that year, a July 2007 Nielsen/Netratings survey of online display advertisers showed that the top five of those advertisers were all involved in the mortgage lending industry to some extent, delivering almost $200 million in monthly revenue to online advertising companies like Google, MSN, and Yahoo! 54 These delivered hundreds of billions of views of online ads that helped drive the frenzy of refinancing and subprime mortgages with ads like the ubiquitous “LowerMyBills” and other online enticements. These numbers are only for display ads online; search advertisers don’t share data on specific revenue from particular companies, but reports at the time showed that mortgage loan companies were paying top dollar for keywords like
“mortgage” and “refinance” with prices going for as much as $20 to $30 each time a user clicked on a search ad.\textsuperscript{55}

Online companies would then sell information about the users identified as likely prospects to mortgage companies, which in turn would contact them. Customers targeted through these online leads for subprime mortgages were disproportionately low income, black and latino. Usually unaware that better deals existed, studies showed that people of color offered these subprime mortgages were 30% more likely to be charged higher interest rates compared to white borrowers with similar credit ratings.\textsuperscript{56} Burdened with unrealistic “teaser rates” that appeared affordable, these loans would explode into unmanageable debt in later years.\textsuperscript{57} This was the most toxic version of price discrimination possible and led to one of the largest scale destructions of wealth among low-income and minority communities in the modern era\textsuperscript{58}, even as the data platforms that helped facilitate this process continued to explode in value and profitability.

Even today, the financial industry remains a bedrock of revenue for advertising-driven big data platforms. According to WordStream, a company specializing in helping companies bid effectively on Google Ads, the three most expensive categories of keyword searches as measured by cost per click are in financial services (insurance, loans and mortgages), with 45.6\% of the top 10,000 advertising keywords falling in those categories.\textsuperscript{59}

Depressingly, bottom-feeding subprime mortgage offers were replaced in the aftermath of the financial crisis by companies exploiting the financial distress of families, particularly by payday loan lenders who offer extremely high-interest loans in exchange for a commitment for repayment from a person’s next paycheck.\textsuperscript{60} Such loans have been banned or severely restricted as exploitative in multiple states and the Consumer Financial Protection Bureau (CFPB) has held hearings specifically on abuses in the industry, with CFPB head Richard Cordray saying “some payday lenders [are] engaged in practices that present immediate risk to consumers and are clearly illegal.”\textsuperscript{61} Their ubiquitous presence in online ads is not an accident; in fact, data platforms have actively solicited ads from the industry, including Google setting up a booth at the annual convention of the “Online Lenders Alliance,” a trade group made up primarily of payday lenders. Industry observers like Robert X. Cringely, who has covered Silicon Valley for over twenty-five years, argue that Google buries bad news about the industry in its search results, “below the fold as we used to say in the newspaper business.”\textsuperscript{62}

Whether or not, as Cringely argues, data platforms do hide negative information about the evils of many of their online financial advertisers, what is true is that they proliferate in the feeds of low-income Internet surfers. As many families saw their mortgages balloon above the value of their homes, an array of illegal scam “loan modification” firms appeared promising to help homeowners in advertisements appearing when people searched for keywords such as “stop foreclosure,” then taking money from those families without helping them at all. Despite scathing reports highlighting the problem by consumer group Consumer Watchdog in 2011,\textsuperscript{63} Google refused to stop until ordered by the Treasury Department using its TARP authority to shut down ads by 85 of the companies. “Many homeowners who fall prey to these scams, initially do so through these Web banners and other Web advertising,” Christy Romero, Deputy Special Inspector General for the Troubled Asset Relief Program (TARP), said in an interview.\textsuperscript{64} Similarly, the data broker and credit score company Equifax kept selling lists of people late in paying their mortgages to fraudulent marketers until the FTC fined Equifax $1.6 million in 2012 for the practice based on companies bilking those customers of millions of dollars.\textsuperscript{65}

In this way, the data and privacy lost by consumers has translated into tens of billions of dollars in profits for the data platforms and the enabling of exploitation by predatory companies using that data for an even larger scale of economic losses by consumers.

### Consumers Lose Financially as the Value of their Personal Information Flows to Big Data Platforms

Beyond losses from price discrimination and from direct scams using targeted data, consumers lose out online as the value of their personal data is coopted for the profits of the big data platforms. In a broad sense, users lose out doubly since the data platforms not only sell their data to
advertisers but also use the free labor and data provided by all users collectively to attract users to their sites in the first place. While users may vaguely feel that they are giving up some control of their data in exchange for services provided by big data platforms like Google and Facebook, those companies depend on the free labor of individuals posting their updates to Facebook, reviews on Amazon, edited stories on Wikipedia, and their links on blogs to make those services valuable.

In fact, Google’s original innovation in search technology was built around harvesting the diffuse labor of people across the Internet. Its original Page Rank algorithm used the links to other websites created by web site creators as a tool to assess and rank the likely value of websites containing similar information or keywords, an algorithm which has only been strengthened by tracking the sites for which people search. Each click adds to the algorithm that can direct users with similar searches and interests to see the same link highly ranked as well. The more people find and use other people’s content via Google, the better Google’s algorithm becomes, reinforcing the precision and strength of its search engine vis a vis any challenger search technology which would lack access to the network of users and the information they generate on search preferences. Similarly, social networks like Facebook and LinkedIn depend on the daily infusion of writing and links by their users to provide value to other users and use experiments on the behavior of those users to strengthen their algorithms.

These big data platforms have positioned themselves to take advantage of what media studies professor Clay Shirky has labeled a profound shift in models of product where “user generated content”, what was once called “free” time,” becomes incredibly valuable when aggregated. That so much seems free on the Internet is just the flip side of people providing so much free labor without being paid themselves and then accessing it on big data platforms. Nicholas Carr has referred to this as “digital sharecropping” where the Internet “provides an incredibly efficient mechanism to harvest the economic value of the free labor provided by the very many and concentrate it into the hands of the very few.” Recognizing the economic gain from user production of information, companies like Facebook go out of their way to encourage the maximum sharing of data possible, using what media activist Cory Doctorow calls “very powerful game-like mechanisms to reward disclosure” with plenty of rewards in the form of “likes and attention from friends and family when they post.”

The incredibly outsized stock valuations of web-based firms such as Amazon (with its user-generated product reviews), Facebook (with its user-generated content and social links) and Google (search for others’ content, user-generated YouTube videos etc.) can best be understood in terms of the free labor and data each is harvesting. For example, when Facebook went public with its initial public offering (IPO), one analyst estimated that users at Facebook had generated 2.1 trillion piece of “monetizable content” between 2009 and 2011, which translated into about $100 billion of the value of its stock market capitalization-- with each Facebook user contributing around $100 of user labor on average to the stock wealth created for Mark Zuckerberg and his fellow shareholders.

Analysts like Michael Fertik, CEO of the company Reputation.com, which helps keep their information anonymous online, estimates that data can be worth in the thousands of dollars each year to all the data platforms a consumer may use. One other measure of the value of user data is the fight Apple had with publishers over terms for sales of subscriptions iTunes. Most publishers were willing to pay 30% of their subscription price to Apple but balked at Apple retaining control of data on subscribers, indicating that publishers valued the user data at more than 30% of the cost of any purchase online.

Yet while publishers negotiate hard over control of that personal data with Apple, the consumers themselves generally give their personal data away for free without a thought. Consumers underestimate the value of their data and lose out
continually in these online transactions. De facto they are in a barter relationship with big data platforms, trading data for access to those “free” services—and the economic history of barter is that less sophisticated partners in such exchanges inevitably lose financially.\(^75\)

Multiple studies show most consumers don’t even understand that their private information used by big data platforms is also being shared with third parties to assist in marketing advertisements.\(^76\) Users rarely read the fine print when they click acceptance of the terms of service on these sites and receive little information about the consequences of sharing their data. Yet this sharing of data with third parties doesn’t reflect consumer preferences: a 2012 Pew survey found 73 percent of the American public were opposed to search engines tracking their search history even to improve search results and 68 percent opposed using user data to help advertisers target ads.\(^77\) Users who understand that such sharing is happening express frustration that they lack the capacity to stop it\(^78\), even though the desire to stop such tracking, aggregating and sharing of data has been increasing.\(^79\)

Notably, the former Federal Trade Commissioner J. Thomas Rosch expressed concern in his 2013 opinion about Google’s potential monopoly abuses that the company engages in “telling ‘half-truths’”—for example, that its gathering of information about the characteristics of a consumer is done solely for the consumer’s benefit, instead of also to maintain a monopoly or near-monopoly position.\(^80\)

Most data platforms declare that a single click signing up for a service creates “consent” by any user for whatever purpose the companies chooses to use their data. The problem is that if few users know how the data is actually being used, such consent is meaningless. And since the companies themselves don’t even know the economic value of user data at the time it is shared, often waiting months or years to figure out how to monetize it, it’s extremely unclear how users can be in a position to effectively negotiate a fair economic value for their data when they aren’t and can’t be told about its potential future use.

### Consumers Lose Access to Local Journalism as Advertising is Diverted to Big Data Companies

When advertising was generally attached only to
specific media products, a reader of a newspaper or magazine helped attract advertising revenue to that outlet every time they read it. This cut the newsstand price and funded journalism staff to cover local and national news across the country.

However, behavioral profiling and tracking users based on their interests, including what people read, means that advertisers can reach the target audiences who read particular newspapers or magazines by advertising through big data platforms instead of buying ads at the magazines or newspapers themselves. Advertisers can pay online advertisers to place ads in lower-cost websites primed to display ads to anyone who previously viewed say the New York Times, reaching those readers without paying a dime to the New York Times itself. This de facto diverts revenue from newspapers and magazines to the big data platforms and undermines the revenue base that previously sustained their production.

As the graph in this section highlights, newspaper-advertising revenue crashed over the last decade by almost exactly the same amount that Internet advertising increased. Between 2005 and 2013, newspaper advertising in the United States went from bringing in a bit more than $47 billion per year to fund newspapers – the largest category of advertising – down to just $18 billion by 2013. In the same period, Internet advertising went from just over $12 billion to roughly $43 billion, with Internet advertising now taking the largest share among media categories.81

The New York Times itself has, with some cutbacks, been able to sustain itself because of the strength of its brand and ability to charge online subscriptions, substituting a higher price charged to its readers to make up for the lost advertising revenue of those same readers that the big data platforms have captured. But for regional newspapers around the country, the result has been the evisceration of newsrooms and even elimination of whole newspapers. One measure is that between 2003 and 2014, there was a 35% decline in full-time reporters covering statehouses across the country.82

Consumers of newspapers lose out since they get a worse product for the same subscription dollars – since those subscription dollars are matched by far fewer advertising dollars. More broadly, society and our democracy lose out as well since we end up with less independent investigations of our government institutions and more opportunity for corporate special interests to pad their profits with tax breaks and other corporate welfare unreported to the public—an indirect way big data platforms also contribute to worsening economic inequality in our society.
III. How Big Data Reinforces Market Power

The deeper problem for consumers and society is that even as big data is damaging consumers in a range of ways, it is also reinforcing the market dominance of increasingly centralized data platforms in various sectors throughout the economy. This dynamic is driven by a combination of so-called network effects, where users group themselves around services already popular with others, how control of unique user information by those data platforms creates a barrier to entry for new competitors, and the ways existing platforms spread into additional sectors to reinforce their market dominance.

Network Effects Lock-in Market Power by Big Data Platforms

Many of these data platforms attract users based on other users participating in those networks, a prime example being social media like Facebook, where consumers become mutually locked-into a service. This is reinforced as users share information with one another, which in turn means there is data on those networks exclusive to those data platforms. Metcalf’s Law, named for networking technology pioneer Brian Metcalf, argues that the value of a network increases as the square of the number of users, which means larger networks are exponentially more valuable than even slightly smaller networks, making it extremely hard for runner-ups to thrive in networked industries.83

Network effects reinforce market power for already dominant companies in a range of ways.

• More Users Attract More Users: Just as the usefulness of telephones exploded in the early part of last century as more people had telephones of their own, so too a social network like Facebook become more useful as a higher proportion of ones’ friends of colleagues are participating. For those looking to reach a broader audience through social networks, participating in the largest networks like Facebook or Twitter give you far greater chances for shares and reshares that take your message “viral” on a larger scale. Or as Niraj Dawar writes at the Harvard Business Review blog, “There is only one square in the global village, and it is run by Facebook. Being on a different square from everyone else doesn’t get you anywhere — you just miss the party.”84

• Dominant Networks Attract Third Party Developer Support: A higher number of consumers of a product or service attracts more third party developers creating other products that work with the core product, which in turn attracts more users to that core product. A prime example is historically how so much software written for the dominant Windows desktop operating system strengthened its desktop dominance, while the number of software apps developed for the Apple mobile operating system and for the Android mobile software deters users from buying phones/tablets in other networks.

• Networks Attract Advertisers: Large networks of users obviously attract advertisers based on the scale of opportunity to reach consumers. However, the dynamics of scarcity in screen space online creates pressure on advertisers to spend on the largest data platforms. With Google, for example, since ads are listed in order based on the most successful bid at auctions for keywords, any company has to worry not only about reaching the audience but at being better positioned on the screen then their nearest competitor. Companies often feel they can’t afford to cede the largest bundle of consumers to the competition so the largest networks get the most bids, driving up advertising prices per user compared to smaller networks.85

Because personal data created by users is often linked to other data within the service or to personal information of other users, it is often hard for consumers to switch to an alternative service without a lot of effort and often losing many of the contextual links, comments by other users and other information embedded in the
architecture of particular data platforms, making any ability to port all personal information to another service nearly impossible.86

Data Platforms’ Control of User Information Creates Barriers to Entry for Potential Competitors Lacking That Data

As a handful of data platforms generate massive amounts of user data, the barriers to entry rise since potential competitors have little data themselves to entice advertisers compared to the incumbents who have both the concentrated processing power and supply of user data to dominate particular sectors.

While control of user data has not been central to traditional debates on competition and antitrust, regulators are increasingly noting the threat of such control to competition. In January 2013, after the FTC dismissed one antitrust approach to Google, Federal Trade Commissioner J. Thomas Rosch criticized his colleagues for not even investigating Google’s collection of user data in its antitrust proceedings, since Google’s “gathering of information about the characteristics of a consumer [may be] done…to maintain a monopoly or near-monopoly position.”87 In Europe Joaquin Almunia, then Vice President of the European Commission responsible for Competition Policy, argued in a speech that “exclusive access to personal data in a given market could give rise to concentration concerns.”88 Germany’s economy minister Sigmar Gabriel has referred to the big data platforms as engaged in “brutal information capitalism” that may be threatening competition.89

The network of users built by big data platforms produce its own network effects as the data produced grows exponentially with the users searching, linking, commenting and producing shared content. Apple made sure that each iPhone sold added to its growing data trove about use of those phones by its users, Facebook and Twitter meticulously tracks sharing across its network, and Google learns more about what users want to find as they track each search and click across their network.

Controlling More Data Increasingly Trumps Other Forms of Innovation

Increasingly, analyzing the maximum quantity of data and finding useful correlations within that data is trumping other forms of insight and innovation, making it even harder for any startup with the proverbial better idea to win out against the pure quantities of data held by incumbents.

One often-cited example of how brute force data analysis beats other forms of insight is the evolution of computer chess-playing computers. Decades ago the assumption was that computers needed to in some way develop a deep and complex analysis of the game to beat human players. Instead, programmers found it was more effective just to feed in massive amounts of information about recorded chess games and let the computer search for past game positions that match whatever game it’s currently playing, then make the move that worked in the past. Similarly, translation software has largely abandoned algorithms that try to understand how language works in favor of feeding in massive amounts of previously translated material (the United Nations is a favorite source for this) where the software just finds a phrase fragment from past translations to help translate a current document, the method for example used by Google Translate to become so successful.90

People don’t always think of Wal-Mart as a big data company, but it pioneered massive data analysis in the retail sector, using extensive analysis of sales data to better coordinate logistics and restocking planning as a competitive advantage. When Amazon launched, it could borrow many of Wal-Mart’s tools (and some of its programmers) to strengthen its own logistics; it then used what users recommended to compare shoppers to other similar shoppers and recommend new purchases. Amazon found human editors recommending products could not begin to match this brute force data analysis and the result was that today, a third of all Amazon sales are from its data-driven recommendation system—and competitors without similar volumes of consumer data have been driven out of business across the country.91

For Google, its network delivers not just data on each individual user, but also the cumulative data
on how similar users behave. This allows Google to anticipate a user’s interests (and thereby increase the returns on advertising) based not just on each user’s own previous actions, but on the behavior of similar users in its network. While Google no doubt got a leg-up into its dominant position through innovation in its core algorithm, even its own leaders have admitted that it has consolidated its control based on overwhelming control of data. As its Chief Scientist Peter Norvig has observed, “We don’t have better algorithms than everyone else; we just have more data.”92

The triumph of raw data analysis means “more trumps less. And sometimes more trumps smarter,” argue Kenneth Cukier and Viktor Mayer-Schönberger in their recent book, The Rise of Big Data: A Revolution That Will Transform How We Live, Work, and Think.93 Building a better mousetrap loses out in the market to companies with the most data.

Mid-size companies don’t benefit much from big data analysis since they usually lack the critical mass of data and can’t invest in the computing power to gain the insights driving market power by the bigger players. Fighting credit card fraud is one example of this; big data analysis tracking likely fraud based on the purchasing patterns of thieves gave larger financial institutions a big advantage in catching such fraud. The result was smaller banks avoided issuing their own credit cards, passing the business onto firms like Capitol One and Bank of America which metasized to dominance of the field. This in turn further strengthened the big banks’ data about consumer finances and their overall market power and dominance in the financial sector.94

**Control of Unique Data Strengthens Control by Big Data Platforms**

As companies collect user data, they gain competitive advantage against any potential challenger who will lack access to that unique user data in setting up any rival service. Such data can be redeplopped by dominant players not just to strengthen their position in existing services but used in related new services to expand their economic reach. In this way, you see Google, Amazon, Apple and Facebook expanding rapidly into a multiplicity of emerging data-related fields, making it extremely hard for upstart companies to get a toehold except in very specific niches.95

There has recently been a flurry of political interest in abusive practices by data brokers who buy and sell personal data, with major reports released by both the U.S. Senate96 and the Federal Trade Commission.97 While the consumer harm detailed in those reports are important, it is worth noting that the companies involved are relative minnows in the big data ecosystem compared to the major big data platforms — and are likely to be even more marginal over time. Experian is one of the largest at $4.8 billion in sales per year98 while Acxiom, a data broker often cited as having one of the largest datasets on consumers, has only about $1 billion per year in revenue.99 Even collectively, these data brokers are dwarfed by a company like Google with over $60 billion in annual revenue.

This is largely due to the fact that most data brokers do not control unique information about individual consumers but instead are merely middlemen. While smart entrepreneurs running such firms were able to position themselves to benefit as corporate advertisers began engaging in targeted advertising, power in the data-driven economy is going to inevitably move to data platforms like Google, Facebook and other companies that continually generate new unique data on consumers. As Cukier and Mayer-Schönberger argued in their book, The Rise of Big Data, the initial skills of data brokers is inevitably losing out to companies “holding large pools of data and being able to capture ever more of it with ease...large data holders will flourish as they gather and store more of the raw material of their business, which they can reuse to create additional value.”100

The key importance of and market power by data platforms controlling unique data is reflected in recent acquisitions. Facebook’s buyout of the photo-sharing company Instagram and of the global texting company WhatsApp were based on gaining control of the massive user base and unique data generated by those users;101 Google has largely focused on growing its own data sources but it’s recent large percentage investment in the taxi service Uber reflects its interest in having a stake in the transit and logistics data being generated by users of that company.102 And the potential scope of Google’s data ambition is reflected in its acquisition of Skybox Imaging, which puts into orbit low-flying
satellites which photograph the whole planet twice a day with the accuracy to count cars in stores’ parking lots and predict those company’s next quarter sales figures, or to survey the health of cropland across earth and estimate the likely price of grain months in advance. In each case, the big data platforms gain access to unique data that leaves potential competitors with even less openings to challenge their market dominance.

**Why Control of User Data Matters: Users are the Product, Not the Consumers, of Advertising-Supported Data Platforms**

While all companies increasingly use big data to strengthen their competitive position, user data is the singular critical element for market power in advertising-supported sectors, because the core product sold to advertisers is access to that user data. (For data platforms like Apple and Amazon whose products are sold directly to users, not advertisers, this dynamic is different). Media studies professor, Siva Vaidhyanathan, has noted that we have allowed our “our fancies, fetishes, predilections, and preferences” to be captured by companies like Google and resold to their advertisers.

Search, email, online video, social media—these aren’t products sold to users. They are offered to users for free essentially as bait to encourage them to give up their data and agree to become a product packaged by behavioral and demographic category to advertisers. And that user product is not merely a compilation of user activities in any one service but of the combined activity across all the services provided by that advertising-based platform. The more services the user uses, the more data to be packaged for the advertising clients.

So when thinking about competition in these advertising-driven data markets, the point is not how easy it is for users to switch to an alternative product but whether advertisers can easily substitute an alternative online advertising service for a dominant one like Google. Competition in search is not the relevant market; competition in search advertising should be the focus for regulators and the Federal Trade Commission itself has argued that search advertising is even a distinct market for antitrust analysis compared to online display advertising, as have competition authorities in other countries. It’s worth emphasizing that the FTC’s ruling in January 2013 was only that Google had no harmful monopoly in search as a market, without any finding or analysis of whether Google’s dominance of search advertising violated antitrust law.

For advertisers seeking the product Google sells, namely ads tailored to targeted behavioral characteristics of online users, there is no economically viable competition to Google that has the same one-stop reach. Facebook and Twitter are working hard to develop alternative sources of user data to get a similar claim, but the competition for online advertising overall is increasingly limited to a narrow set of players— and search advertising has largely been ceded to Google alone. The people who use Google search are the product sold to those search advertisers and to the degree Google has unique data on those users, any advertisers wanting to reach those users has to go through Google. And because no new competitor can get access to that user data without getting the scale of Google-- and can’t get that scale without the user data -- you have little chance of new competition displacing Google’s dominance in search advertising. Even Facebook is making inroads only in the much smaller sector of online display advertising online and mostly at the expense of other players, not Google.

Evidence of this market power from controlling user data is the premium price Google is able to demand from advertisers, one reason that the company is able to translate 65% of all searches by United States users online into 78% of all search advertising revenue. Yet while a dominant player like Google is inevitably going to deliver more clicks on any keyword because it has more users, all things being equal, the Cost Per Click (CPC) price – how much an advertiser pays when an online user clicks on an ad -- should be roughly the same since a user ultimately clicking through to an advertiser’s page should in theory be just as valuable if the customer reaches the page via Google, as via AOL or via Microsoft’s Bing.

However, even in the United States where Bing has a significant minority share of searches— especially when including its alliance with Yahoo!—Google receives a high premium on its CPC. One advertising analyst estimated that for
the same keywords, the “average CPC on Bing is somewhere around 1/4 or 1/5 of our average CPC on Google.” Another found that on specific search terms, CPC rates on Bing were slightly higher but still were discounted 49% to 71% compared to Google.

What this means is that even if a United States advertiser pays for a search term on both Google and Bing, Bing would end up generating only one-tenth to one-fourth of the revenue Google received from the same advertiser’s campaign using the exact same keywords on each site (i.e. roughly half the number of clicks times one fifth to half the revenue per click). This explains at least part of the reason why Bing may have nearly half the U.S. users of Google, but generates less than twenty percent of sector revenue.

The lower CPC rate means that potential competitors to Google receive much less revenue but have to invest similar amounts in fixed costs as Google to maintain a competitive search engine and platform. Bing has not been able to attain a financially viable competitive challenge and the fixed costs for entry create a major barrier to entry for any new potential Challengers to Google, so its market dominance seems unlikely to be eroded by market forces.

As Jonathan Rosenberg, Google’s own VP of Product Management and Marketing explained in an unguarded statement in 2008:

> We get more users because we have more advertisers because we can buy distribution on sites that understand that our search engine monetizes better. So more users more information, more information more users, more advertisers more users, it’s a beautiful thing, lather, rinse, repeat, that’s what I do for a living. So that’s ... the engine that can’t be stopped.

That Google’s own executives argue that its control of user data has helped create a self-reinforcing circle that has strengthened its dominance should encourage regulators to investigate how Google and other big data platforms use user data to create and extend their market power.

Aside from the broader harm to society and consumers outlined earlier in this paper, the most basic consumer harm that should demand a regulatory response are the inevitable higher costs to the advertising customers. Now, average people don’t necessarily have to cry tears for those advertisers, but from a competition viewpoint, higher costs spilling into the economy due to monopoly is a sufficient basis for antitrust action. There are plenty of antitrust cases over the years over higher costs charged to advertisers, one reason the government had to grant special antitrust dispensation in the Newspaper Preservation Act to many of the joint advertising operations by newspapers. Newspaper advertising was considered a unique sector and other forms of advertising were not a legal substitute, so dominance in reaching any unique set of readers was anti-competitive. Similarly, the Federal Trade Commission itself has affirmed that search advertising is a different industry sector for analysis from online display advertising.

What makes the harm to average consumers from current dominance by big data platforms seem distinctly worse is that the higher profit margins for those data platforms do not seem to be plowed back into providing equally valuable services for end users in the same way advertising in traditional media was reinvested into media content. What Google users get in free services is a tiny fraction of the value of the data they have given up and the advertising profits generated by Google. Instead, Google has plowed that revenue into new ventures -- buying YouTube, expanding Android, moonshot projects like driverless cars, Google Fiber, etc. -- and just piling up its current $59 billion cash on hand. Users aren’t getting the value for their data, either in cash payments or in services provided, which goes to the consumer harm that differs from other traditional advertising-supported media where the value provided to readers/viewers was pretty comparable to the advertising dollars generated.

### Expanding into Multiple Related Sectors Reinforces Market Power in Core Data Platform Businesses

By spreading into multiple related data-driven business sectors, the dominant data platforms reinforce market power in their core businesses. Expansion into new sectors adds to their overall storehouse of user data, while choking off other sources of data for potential competitors.
Google has been the most aggressive player in this, expanding from its core search advertising business into almost every imaginable space where users operate online (and increasingly offline), from watching videos, to emailing friends, to buying products, to using their cell phones, to updating their calendars. Google’s reach into all these activities allows it to develop an integrated profile of more individuals with greater breadth and depth than any potential rival.

The promotion of free products online becomes what analyst Bill Gurley calls the creation of “moats” around Google’s “economic castle:” online search advertising. Gurley argues:

Android, as well as Chrome and Chrome OS for that matter, are not “products” in the classic business sense. They have no plan to become their own “economic castles.” Rather they are very expensive and very aggressive “moats,” funded by the height and magnitude of Google’s castle... Google is also scorching the earth for 250 miles around the outside of the castle to ensure no one can approach it.113

Moving into these new online sectors is therefore less to create new revenue-generating ventures than to scoop up new data and, as importantly, prevent potential rivals from gaining a critical mass of user data where they might challenge where a company is dominant, as with Google in search advertising. This is analogous to how the U.S. Court of Appeals in 2001 described Microsoft’s actions where that company’s expansion into web browsers and other so-called online “middleware” products was less about dominating those Internet sectors than “to meet the threat to Microsoft’s monopoly in another market (operating systems).”114

Google has moved heavily into e-commerce, from launching a Google Advisor site115 to offer mortgages, credit cards and other financial services from partner companies to expanding its Google Shopping service where companies pay fees to be listed in shopping searches116 to its recent launch of Google Shopping Express promising same day delivery of products from partners like Target and Costco.117 The goal for Google is less to compete with Amazon in revenue from such sales (since every indication is that Google runs these e-commerce programs with deep deficits) than to capture a far deeper profile of user purchasing data. Every click of a button at a checkout counter through a Google shopping program delivers data that can help Google track nearly the whole lifecycle of financial behavior from what goods and terms people use to search for products to what ads they click on associated with those search terms to which offers they download to where they then make a purchase and redeem any offers.118

The battle for control of the mobile phone and tablet market highlights the dynamics of this new kind of data-driven fight, where companies enter various markets less to generate revenue in that sector than to secure dominance in its existing area of market power. The phone/table market has four behemoths – Google, Apple, Amazon and Microsoft – competing for control, yet Google pushes out Android phones and tablets mostly to maximize its collection of user data, particularly location data, rather than to maximize profits from device sales; Amazon is involved in selling tablets at a loss or near breakeven point largely to drive users to its e-commerce offerings;119 Microsoft entered the market for the same reason it bothers to compete at a loss with Google in search engines—to protect its Windows operating system and productivity software as a standard for home and business use; only Apple’s business model is primarily based on actually making profits on its revenue from tablet and phone sales directly.

Google as a Focus for Initial Action Among Big Data Platforms

While the danger to consumers and the prospect of increasing economic inequality discussed in this paper is reflected in dynamics across the online (and increasingly offline) economy, Google could be a key initial target for advocates and regulators to focus on since it is in many ways the leading edge of a broader set of data platforms that are all increasingly consolidating dominance across different economic sectors.

Currently, no other data platform dominates its sector quite as completely as Google dominates the search-advertising sector. While platforms like Apple and Amazon use big data to strengthen their market power, the fact that their revenue model is based on selling products directly to consumers...
means they do not engage quite as much in the privacy-destroying practices that Google does as an advertising-oriented company serving third-party business customers.

Google’s dominance of the search advertising and related sectors is quite comprehensive. There are over billion people globally using Google products: search engines, YouTube, Google Maps, Android mobile devices, e-commerce offerings and so on, all feeding data into and generating ads and profits for the company.

Google is estimated to have increased its global share of search advertising from 72% in 2006 to over 85% by 2012 (and 78% of the U.S. market alone). In mobile phone search advertising, the fastest growing online advertising sector, Google has established a phenomenal 90% of mobile search advertising, which was projected to be worth $8.8 billion in 2013 for Google and is likely to skyrocket as a percentage of all online advertising.

Overall, Google reported annualized revenue of well over $66 billion in annual revenue in 2014—a 19% increase from a year earlier—with annualized profits of over $19 billion per year. Google is essentially the only company making any profits in the search-advertising sector. Yahoo, once one of Google’s main competitors, has abandoned search in favor of an alliance where Microsoft’s Bing search engine powers most search on Yahoo properties—and its payments to Microsoft largely equaled its costs by 2011. And Bing has been reported to be losing something on the order of $2.6 billion per year for Microsoft. That even a deep-pocketed challenger such as Microsoft cannot create a viable challenger to Google’s search advertising dominance highlights, as will be explored later in this paper, the way the dynamics of big data platforms reinforces market power more generally.

Lack of High-Tech Competition
Highlights Need for Action by Regulators

The heavy competition in phones and tablets is increasingly the exception and related to the fact that most of the companies competing in that sector don’t see it as a primary profit-center; in those core profit centers, companies increasingly seek and use the power of big data to gain as exclusive dominance as possible. Where that dominance does not come organically, companies can just acquire any startup likely to infringe on that market power. For example, Facebook extended its social media presence first by its billion-dollar purchase of photosharing company Instagram and, more recently, its $19 billion takeover of WhatsApp, one of the largest global texting platforms, essentially eliminating any significant rival in the various iterations of global social networking.

The combination of incumbent control of unique data, the willingness of the behemoths to run at a loss in multiple related sectors in order to reinforce power in their core profit-making “castles,” and their increasing willingness to buy out any potential rival long before they become a viable competitor have all contributed to the broad decline in competitiveness and entrepreneurship in high-tech markets.

Contra arguments that antitrust and other pro-competition regulation is outmoded in the new, high-tech economy, the reality is that the rise of the Internet has coincided with a drastic decline in new competitors emerging to challenge dominant firms. Reflecting the dominance in these silos, most new small companies aren’t looking to go public with an initial public offering (IPO), but instead are “cashing out” by getting acquired by existing big firms like Google, Amazon or Microsoft.

Where there were an average of 311 IPOs per year during the 1980-2000 era, researchers Xiaohui Gao, Jay R. Ritter and, Zhongyan Zhu found there were only 102 IPOs per year during 2001-2009; that’s a decline of two-thirds and reflects what those researchers see as an economy where minnows don’t grow up to challenge the big fish but are being eaten to fatten already existing dominant, often nearly monopolistic firms. This is reflected in this graphic on the next page courtesy of Zero Hedge about the long-term collapse of startup firms going public with IPOs.
For smaller companies looking to go public, the IPO market has been even less active. In 1999, there were 249 IPOs that raised between $5 million and $60 million from the market. In 2012, there were just 12.\textsuperscript{129} In an economy of super profits for big data platforms and close to non-existent profits for smaller firms, it makes sense that most firms are looking to cash out through acquisition. Snapchat may have made news by turning down a big buyout offer from Facebook to remain independent\textsuperscript{130}, but it’s news precisely because of how unusual that choice has become, despite the handful of high-profile IPOs like Facebook, Twitter and LinkedIn.

The reality is that few new firms are joining the ranks of top companies in the economy. A 2010 Kaufman Foundation study found that four-fifths of the Fortune 500 came into existence before 1970, over forty years ago. Firms created during the 1910s a century ago account for more firms in the Fortune 500 than firms created in either of the last two decades of the "Internet economy."\textsuperscript{131} Similarly, a Brookings study by Ian Hathaway and Robert Litan found that the share of companies that are more than 15 years old has increased dramatically since the early 1990s.\textsuperscript{132} Given that the market is producing fewer likely challengers in the marketplace to keep dominant data platforms in check, it becomes even more important for regulators and courts to take action earlier in technology markets, before market dominance arises. The standard worry about antitrust (or any other government regulation) is that it will undermine innovation, yet centralized power in a sector is just as likely to undermine innovation by the leader discouraging innovation by potential entrants to a sector. Greater regulatory intervention could actually encourage competition that better serves user privacy and consumer economic interests.
IV. Taking Action to Protect Workers, Consumers and Society in an Era of Big Data Platforms

The upshot of this market power by big data platforms is that the marketplace is doing little to create options for consumers that might alleviate the misuse of consumer data or encourage big data platforms to better compensate users who are willing to share their data. As the accumulation of all this data is in increasingly fewer corporate hands with little market pressure on those companies to respect the privacy of users, economic justice advocates should be engaging regulators to take action to prevent those big data platforms from using that data in ways that harm workers and consumers, particularly low-income, minority and other vulnerable members of the population.

Promote “Self-Help” Campaigns to Empower Consumers to Take Control of their Own Data

A first step for advocates may be to strengthen and expand on ongoing campaigns that seek to rein in data platforms through boycotts, promoting “self-help” technological tools to protect consumer privacy.

Various privacy groups such as the Electronic Frontier have campaigned to pressure companies creating web browsers and other online software to include tools such as “do not track” options that prevent advertisers from following users across the web. A number of other campaigns have directly tried to shame platforms for carrying ads or otherwise collaborating with noxious or predatory advertisers, such as the campaign discussed earlier by Consumer Watchdog to stop Google from carrying scam “mortgage adjustment” ads or pro-choice groups which demanded online advertisers stop carrying fake “abortion care” ads from anti-abortion centers. Expanding these kinds of campaigns could help educate the public about how they are being tracked online and the ways their data are used to deliver often deceptive and exploitive ads online.

However, encouraging “self-help” actions by consumers is not likely to make a significant difference in actually strengthening consumer power over their data. While a number of companies offer technological tools to supposedly stop tracking by advertisers or increase the anonymity of users online, they largely fail in the face of determined tracking of users by online companies. Julia Angwin details a year-long quest to evade online trackers in her book *Dragnet Nation* and after testing many of the best of the technological tools available (with the tech support available to a top reporter at the *Wall Street Journal*), her conclusion was that it was largely a hopeless enterprise.

Other analysts like Cukier and Mayer-Schönberger agree: “In the era of big data, the three core strategies long used to ensure privacy—individual notice and consent, opting out, and anonymization—have lost much of their effectiveness.” Given the amount of personal data already out there, even anonymous users are reidentified relatively easily by comparing the language or stray information posted under any pseudonym to information already known about individual consumers. And even where a user abandons all use of “cookies” online, advertisers can “fingerprint” their web browser by cataloging the unique configuration of plugins, settings and other features of a browser. Especially for low-income families with less technological savvy, such solutions are almost completely useless.

Recognizing the limits of individual action, some groups argue for creating online consumer coops to strengthen the ability of users to collectively negotiate better terms for use of data by big data platforms. Promoting such coops may be a step in the right direction – and if nothing else could be a vehicle for supporting a range of other campaigns – but will be unlikely to gain the traction in the absence of broader regulatory and legal changes in consumer rights.

Limit Employer Profiling in Hiring and of Current Employees

Employers can too easily abuse data they can obtain on potential or current employees not to hire the best worker but to screen out those with the “wrong” opinions or other characteristics irrelevant to the skills need for the job. Most dangerously, people likely to demand better treatment in the workplace or who might join a
union can be screened out, lowering wages in that workplace and, as such practices are compounded across most workplaces, systematically increase economic inequality.

Policymakers should ban or restrict employers from using any data except that most relevant to the position a company is hiring for in order to ensure both employee privacy and that people aren’t punished for free speech and freedom of association before they ever get a job in the first place.

Strengthen Regulations and Laws Protecting Consumer Privacy

In March 2012, the Federal Communication Commission (FTC) issued a report, Protecting Consumer Privacy in an Era of Rapid Change, that sought to outline a framework for privacy protection for both businesses to adopt voluntarily and, where necessary, policymakers could mandate as part of general consumer protection. The framework includes so-called “Do Not Track” rules for web browsers such as Google’s Chrome browsers to ensure user activity can be hidden from advertisers, data portability to allow users to switch easily between email and social networking services and take their data with them, and greater transparency and choice by consumers on where and how they share their data with companies.

However, the law will need to be changed to create a real regime of regulatory and court enforcement against companies violating such anti-tracking and anonymity rules protecting consumers. Up to now the courts have largely failed to recognize a private harm from disclosure of user data to third parties, arguing that since consumers can’t show economic losses from the sale of their data, they can’t seek damages. Possibly as the economic harms of behavioral targeting by online advertisers has become a bit more understood, there have been a few glimmers of movement in the courts, such as a recent 9th Circuit Court of Appeals decision allowing a lawsuit against Facebook to proceed since consumers could have suffered damages due to “losing the sales value of that information” and by its general dissemination. Still, new laws or regulations need to make clear that violation of laws protecting user data create a private right of action by users with clear default economic damages for violations.

There is actually a fairly objective criterion to measure harm to consumers from loss of privacy, even if measurement is not simple. Jaron Lanier framed this as the data platforms should only make money on the value-added that they bring to data collection, but should not be making any money from actually having the data in the first place, since that value should have gone to the individuals associated with the data.

The FTC framework also suggests companies be required to obtain “express content” when collecting “sensitive data,” such as health and other data regulators might deem most subject to abuse. Given the use of a wide range of data for profiling of the financially vulnerable online, such explicit consent should extend to almost all data collected by data platforms beyond the most basic, publicly accessible data about individuals. Detailed and explicit “opt-in” consent should be required for any use of user data with specific express consent required for any change or new use of the data in the future.

While data platforms may express worries that such consent rules will deter use of their services, the very reluctance of consumers to invest the time to complete the process of giving such consent would actually serve a positive purpose in encouraging big data platforms to create economic incentives for users to do so. Tight “opt-in” consent requirements create a “friction” point where people stop and may negotiate a better price for their data. By jumpstarting a real market for user data, it would open up more space for new companies to compete on incentives at that point of friction and potentially encourage all data platforms to either better protect privacy or share some of the profits of the industry directly with users. Limiting such an opt-in requirement for sharing data to larger, dominant players could avoid the problem that general opt-in requirements might lead to users favoring large players to avoid the transaction costs of dealing with multiple, smaller players for their online needs.

One other way to address the fundamental information asymmetry between big data platforms and users in pricing the value of user data would be to adopt proposals that would require greater transparency in how companies...
monetize that data, such as regular reports on the Cost Per Click or other payments companies receive based on user activities. Such information, along with greater data portability between services, might help further a market where users “vote with their feet” (or, more accurately, their data) and demand either a greater share of big data platform profits based on that data, switch to competing providers for a better deal, or withhold their data altogether after recognizing the pervasive use by third parties that they may not want tracking them. Any of those outcomes would lessen the consumer harm from that big data platform control of user data.

Restructure Markets to Promote Competition

However, given the lack of options for consumers in many online industry sectors where one or only a few companies dominate, depending on empowering consumers is only a limited tool. As Julia Angwin notes, helping put a price on personal data only goes so far: “Before we had a minimum wage and limited work hours, people were willing to ‘sell’ their labor at extremely low prices for very long hours.”

There is a strong argument that federal regulators should be looking at how to structure data platform sectors to both promote more competition and encourage more consumer power within those sectors. Given the near-monopoly of certain platforms in search advertising and social networking, European regulators are already looking closely at the connection between industry concentration and the loss of consumer privacy. France’s economy minister Arnaud Montebourg has argued that regulators should be looking at tighter regulation, including potentially moving to “unbundling” companies like Google, for example by separating its search arm, mobile, YouTube and email services into separate companies.

State attorneys general could be approached (with the goal of enlisting the US FTC or Justice Department in the long term) to support an antitrust lawsuit against Google around its overall dominance of the search-advertising sector. There is no exact equivalent in antitrust law precedent to Google’s control of user data but the fact that no potential competitor can obtain enough user data to be a threat to the company’s key search advertising market is analogous to other monopolists’ attempts to deny key inputs to rivals, from the denial of key metals to industrial concerns or, possibly of more relevance, to the illegal attempt by Microsoft to eliminate the nascent threat of Internet browsers to its operating system monopoly by choking off distribution by Netscape. As the D.C. Circuit argued in United States v. Microsoft, courts can infer antitrust liability “when exclusionary conduct is aimed at producers of nascent competitive technologies as well as when it is aimed at producers of established substitutes.”

Given the premium Cost Per Click (CPC) Google can charge advertisers because of its superiority in targeting users, combined with the fixed costs in maintaining such a multi-product platform to access this breadth of user data, it clear that Google will not face viable competition in the search advertising field without antitrust action by government agencies, especially given its relentless expansion into new sectors to expand its control of user data.

Directly Regulate and Restrict Practices that Harm Consumers

On the other hand, encouraging more competition for data platforms and empowering individual users to control their data will still fail to protect many consumers, especially in areas of consumer harm where the danger is indirect, in the future, or deliberately concealed. Because many areas of misuse such as price discrimination, racial profiling and many of the other systematic categorical profiling harms are not obvious when the data is shared and not even fully visible to the people effected by those harms, if policymakers depend only on transparency and consent they will inherently fail to address known harms. On one level, a range of transparency and consent rules covered subprime mortgage borrowers, yet mass scale discrimination and exploitation was widespread. Better transparency can help but when the harm is companies using data to find people who ignore proffered information (Varian’s “myopic” customers), transparency and consent has to fail.

Federal legislators and regulators should engage in direct rule-making to protect basic areas of privacy and to disallow third party use of data altogether where likely victimization of various classes of consumers is likely. As scholar Frank
Pasquale has argued, putting the burden on consumers to discover opportunistic behavior by data platforms is less valuable in many situations than a “collective commitment to privacy.”

One clear step would be to bar data platforms from engaging in price discrimination or from knowingly facilitating price discrimination where different groups are secretly offered different prices by its advertisers for the exact same product or service. As Joseph Stiglitz and Steven Salop argue, government can economize on wasteful information costs in the economy by “eliminating the price dispersion” associated with price discrimination.

One broad approach would be to bring the participation of big data platforms in marketing financial services under the regulation of the Consumer Financial Protection Bureau, which should regularly audit their practices to ensure they are not facilitating predatory price discrimination or other financial scams online. The CFPB is tasked not only with regulating abuses by the banking industry, but it is also required to restrain abuses by “larger nonbank participants” in the financial system. Precisely because so many of these predatory offers are hidden from public view, the CFPB could play a prime role in improving data collection and better assessing the financial harm to consumers from these advertiser practices online. By closely overseeing how online advertising players collect and share the personal data they control with financial services firms, many of the abuses that fueled the concern that created the CFPB in the first place could be reined in before consumers fall victim to fraudulent or discriminatory offers.
V. Conclusion: Stemming Rising Economic Inequality by Better Regulation of Big Data Platforms

As more of the economy moves online, the importance of data mining and the asymmetry of control of information becomes ever more critical in economic markets. Addressing this change calls for far more active regulatory action to reverse the trends undermining user privacy and increasing economic inequality due to that rising information asymmetry. Such action should lead to a greater focus on requiring big data platforms to share the financial bounty of user information with those users, serving both equity and competition.

Data mining of individual privacy is fundamentally reshaping markets by transferring so much knowledge about user interests, behavior and desires into a few corporate hands. Such information asymmetry is easily converted into economic inequality when one side of every transaction has so much more knowledge about the other during bargaining. Joseph Stiglitz and allied economic thinkers have argued that increasing information asymmetry feeds that increasing economic inequality in ways such that the "result from the new information economics is that issues of efficiency and equity cannot easily be delinked." When information is itself the commodity being sold, companies inevitably "appropriate the returns to creating information for economic advantage in the market place" in ways that creates dynamics towards monopoly control of such markets.

Big data platforms have the economic incentive, pushed by advertisers seeking their own information advantage in marketing to users, to increasingly violate user privacy to appropriate more and more information to solidify its economic dominance, all while making it less economically possible for potential competitors to challenge the company. The fact that many of the largest online advertising customers in the mid-part of the last decade were linked to the subprime mortgage industry is just one indicator that understanding the dynamics of the search advertising sector gives insight into larger theoretical problems of market failure, the harm from predatory firms and why we have seen rising economic inequality and corporate concentration in the economy over recent decades.

Government authorities using regulatory tools can stem at least part of this trend by restoring a degree of control by individuals over what personal data is shared online and the financial terms on which that data is shared. The less big data platforms are able to use privacy violations for anti-competitive purposes, the better guardian of legitimate privacy concerns they will become. At least one writer has compared the market failure of providing privacy and data protection to that of poor user information on food and safety a hundred years ago, explicitly highlighting the way equity and consumer safety concerns of crusaders like Upton Sinclair then should be the precedent for action today.

This in turn can eliminate some of the information-based inequality in the modern marketplace that is driving overall economic inequality. If nothing else, public advocacy for these changes can be a chance for a much broader public debate on the abuses of data mining online and how to make all markets work more fairly for average working families.
Endnotes

1 In 2011, the World Economic Forum declared that personal data was emerging as a "New Asset Class." 
   http://www.weforum.org/reports/personal-data-emergence-new-asset-class
2 Jaron Lanier, Who Owns the Future? (2014)
3 Interactive Advertising Bureau (IAB), IAB internet advertising revenue report 2013 full year results, April 2014; 
   http://tinyurl.com/42vo62
4 Ibid.
6 Data Brokers: A Call for Transparency and Accountability (FTC May 2014), http://1.usa.gov/1kbUlCF; Protecting 
   Consumer Privacy in an Era of Rapid Change: Recommendations For Businesses and Policymakers (FTC March 2012), 
   http://tinyurl.com/nwadlb8
7 "Snapchat Settles FTC Charges That Promises of Disappearing Messages Were False," FTC, May 8, 2014, 
   http://1.usa.gov/1kSeVgD; “Facebook Settles FTC Charges That It Deceived Consumers By Failing To Keep Privacy 
   Promises,” FTC, November 29, 2011, http://1.usa.gov/1dwNX0x; “Google Will Pay $22.5 Million to Settle FTC Charges it 
   Misrepresented Privacy Assurances to Users of Apple’s Safari Internet Browser,” FTC, August 9, 2012, 
   http://1.usa.gov/19VRGcD
8 Ki Mae Heussner, “Teacher Loses Job After Commenting About Students, Parents on Facebook,” abcnews.com, Aug. 19, 
10 "Facebook costing 16-34s jobs in tough economic climate," On Device Research, May 29, 2013; 
   http://tinyurl.com/nnn6mmv
11 Kathleen Pender. “Scholarship providers vet students’ social networks.” San Francisco Chronicle, March 6, 2012; 
   http://tinyurl.com/qy5asx
13 Alex Rosenblat, Kate Wiklelius, Danah Boyd, Seeta Peña Gangadharan, and Corrine Yu. Data & Civil Rights: Employment 
   Primer. Data & Society/OIT/Leadership Conference, October 30, 2014; http://tinyurl.com/6t6c5mk
   http://tinyurl.com/nz2j7mv
   http://tinyurl.com/qzpfr8v
16 Kenneth Cukier and Viktor Mayer-Schönberger. Big Data: A Revolution That Will Transform How We Live, Work, and 
   Think (2013), Chapter 4, subsection “Predictions and predilections”
21 “The sale of (anonymous) wireless users’ location and behavior is already big business.” FireceWireless, May 23, 2013, 
   http://tinyurl.com/m9km6x7
23 “Facebook to Partner With Axiom, Epsilon to Match Store Purchases With User Profiles,” Feb. 22, 2013, AdAge; 
   http://tinyurl.com/b2xbzb6
24 For more on specific ways advertisers can target ads based on users behavior, see these Google Help pages: Where ads 
   17, 2014) and Reaching your audience on websites, Google, http://tinyurl.com/pym529m
25 “The quality of the advertising site means whether users find the link useful and are likely to click on it, since Google wants to maximize not just the bid price but its total revenue, including the total number of clicks times that bid price.” George Michie, “Understanding Google’s Latest Landing Page Quality Score Release,” Search Engine Land, Oct. 24, 2011, http://tinyurl.com/3kq68wg

26 Interactive Advertising Bureau (IAB), IAB internet advertising revenue report 2013 full year results, April 2014; http://tinyurl.com/n42vo62

27 Ibid.


29 See Ian Ayres, Super Crunchers, p. 173 (2007) (“[Firms] are becoming more adept at figuring out how much pricing pain individual consumers are willing to endure and still come back for more.”)


32 Angwin (2014).


40 Rosa-Branc Esteves and Joana Resende, Competitive Targeted Advertising with Price Discrimination, Working Paper, March 2011, http://tinyurl.com/p9rxwqw (“price discrimination through targeted advertising may be detrimental to social welfare since it boosts industry profits at the expense of consumer surplus.”)


43 Andrews Ibid., p. 36.

44 James O'Toole, “Wells Fargo in $175M discriminatory lending settlement”, CNNMoney.com, July 12, 2012; http://tinyurl.com/o6dsdu4


49 FTC Data Brokers report, p. 25; http://tinyurl.com/p8jmspuh
systems have collided.

Polonetsky, director of the Future of Privacy Forum, a Washington, D.C. http://tinyurl.com/7kud6fp

("Their raw materials.")

the lion's share of wealth now flows to those who aggregate and route those offerings, rather than those who provide the

Lanier, Vaidhyanathan, p. 30 ("Google is taking a free ride on the creative content of billions of content creators."); see also

Carr, "Sharecropping the long tail," Rough Type, December 19, 2006; http://tinyurl.com/yic5d9c; see also Vaidhyanathan, p. 30 ("Google is taking a free ride on the creative content of billions of content creators."); see also Lanier, Who Will Own the Future? ("An amazing number of people offer an amazing amount of value over networks. But the lion’s share of wealth now flows to those who aggregate and route those offerings, rather than those who provide the “raw materials.”)


Cuiker and Mayer-Schonberger.

Quentin Fottrell, "Who Would Pay $5,000 to Use Google? (You)", SmartMoney.com, Jan 25, 2012; http://tinyurl.com/7kd6f6p ("Their entire market cap is related to how much data is being collected and used,' says Jules Polonetsky, director of the Future of Privacy Forum, a Washington, D.C.-based think-tank.").

Lauren Indvik, "Financial Times Refuses To Give Up Subscriber Data to Apple," Mashable, Apr. 4, 2011; http://tinyurl.com/3ocu44

Karl Polanyi has been one of the foremost documenters of the historical disruptions where market and non-market systems have collided. See The Great Transformation (1944) and Trade and Market in the Early Empires (1957).

95 Id. Just 38% of internet users say they are generally aware of ways they themselves can limit how much information about them is collected by a website.


99 See David A. Balto, "Networks and Exclusivity: Antitrust Analysis to Promote Network Competition," *7 Geo. Mason L. Rev. 523, 524* (1999) (addressing emerging issues involving exclusivity and networks in antitrust analysis and noting that "what distinguishes a network from most other production facilities is that it also enjoys economies of scale or scope in demand"); Steven C. Salop & R. Craig Romaine, "Preserving Monopoly: Economic Analysis, Legal Standards, and Microsoft," *7 Geo. Mason L. Rev. 617, 618* (1999) ("finding of monopoly power is not implausible in a market subject to scale economies and strong network effects").

100 Lanier, *Who Will Own the Future?* ("Even if you capture every little thing you had uploaded, you can't save it in the context of other interactions with other people. You have to lose a part of yourself to leave Facebook once you become an avid user.")


102 Address at Privacy Platform Event: Competition and Personal Data Protection (Nov. 26, 2012), [available at](http://tinyurl.com/nwadlb8)


104 Cukier and Mayer-Schönberger, p. 36.

105 Id., p. 52.

106 Matt Asay, "Tim O'Reilly: 'Whole Web' is the OS of the future," *CNET* (Mar. 18, 2010), [http://cnet.co/ck9ov](http://cnet.co/ck9ov). Similarly, former Google CEO Eric Schmidt has said, "Scale is the key. We just have so much scale in terms of the data we can bring to bear." "How Google Plans to Stay Ahead in Search," *Bloomberg Businessweek* (Oct. 02, 2009), [http://buswk.co/iarA6c](http://buswk.co/iarA6c).

107 Cukier and Mayer-Schönberger, p. 36.

108 Cukier and Mayer-Schönberger, p. 127


105 See Google/DoubleClick, No. 071-0170 at 3 (F.T.C. 2007) (statement of the Commission), available at http://1.usa.gov/RGkEVe (The FTC approved Google’s acquisition of the online display advertising company DoubleClick because “advertising space sold by search engines is not a substitute for space sold directly or indirectly by publishers [of display advertising] or vice versa.”); Press Release, U.S. Department of Justice, Yahoo! Inc. and Google Inc. Abandon Their Advertising Agreement: Resolves Justice Department’s Antitrust Concerns, Competition Is Preserved in Markets for Internet Search Advertising (Nov. 5, 2008), available at http://1.usa.gov/eILMV6 (barring Google from coordinating its search advertising with Yahoo).


107 Natalia Klishina, “PPC Search: Google AdWords vs. Microsoft AdCenter (Bing),” CallFire (Jan. 7, 2011), http://tinyurl.com/3ee57hp


110 See for example Times-Picayune Pub. Co. v. United States, 345 U.S. 594, 610-14 (1953) {Supreme Court distinguished newspaper advertising from “advertising placed through other communication media” based on the “[advertising] trade’s own characterization of the products involved” and the way the advertising industry differentiates “between advertising in newspapers and in other mass media.”}


112 See Google/DoubleClick, No. 071-0170 at 3 (F.T.C. 2007) (statement of the Commission), available at http://1.usa.gov/RGkEVe (The FTC approved Google’s acquisition of the online display advertising company DoubleClick because “advertising space sold by search engines is not a substitute for space sold directly or indirectly by publishers [of display advertising] or vice versa.”); Press Release, U.S. Department of Justice, Yahoo! Inc. and Google Inc. Abandon Their Advertising Agreement: Resolves Justice Department’s Antitrust Concerns, Competition Is Preserved in Markets for Internet Search Advertising (Nov. 5, 2008), available at http://1.usa.gov/eILMV6 (barring Google from coordinating its search advertising with Yahoo).


114 United States v. Microsoft Co., 253 F.3d 34, 60 (D.C. Cir. 2001)
Eclining business dynamism hathaway litan.pdf

See more in discussion by the author at “Google Without Links: Could It Swallow E-Commerce?”, Huffington Post, June 8, 2011, http://huff.to/1qX43rz


See U.S. Senate Data Broker report, p. 3 - http://op.bna.com/der.nsf/id/sbay-9ehtxt/$File/Rockefeller%20report%20on%20data%20brokers.pdf (“consumers generally do not have the right to control what personal information is collected, maintained, used, and shared about them.”)

In Re Facebook Privacy Litigation, NO. 12-15619, May 8, 2014; http://www.leagle.com/decision/in%20FCO%2020140508120

FTC Protecting Consumer Privacy report. at viii.


Alexander Furnas, “It’s Not All About You: What Privacy Advocates Don’t Get About Data Tracking on the Web,” The Atlantic, Mar 15, 2012, http://bit.ly/OX68ld (“The data collectors have more information than those they are they are collecting the data from; the persuaders more power than the persuaded.”)

Whittington & Hoofnagle at 1367


Chris Jay Hoofnagle, Ashkan Soltani, Nathaniel Good, Dietrich J. Wambach, and Mika D. Ayenson, “Behavioral Advertising: The Offer You Cannot Refuse,” 6 Harvard Law & Policy Review 273 (2012) ("We empirically demonstrate that advertisers are making it impossible to avoid online tracking. Study looks at online tracking of cookies and ways companies evade user decisions to refuse tracking")

Frank Pasquale, Beyond Innovation and Competition, 104 Nw. U.L. Rev. 105, 154 (2010)

Salop and Stiglitz at 494.

The CFPB in developing its rules noted the source of this authority. See Defining Larger Participants in Certain Consumer Financial Product and Service Markets, Proposed Rule by the Consumer Financial Protection Bureau, Feb. 17, 2012, Footnote 3; https://www.federalregister.gov/articles/2012/02/17/2012-3775/defining-larger-participants-in-certain-consumer-financial-product-and-service-markets#footnote-3 (“Section 1024 of the Act applies to nondepository (nonbank) covered persons and expressly excludes from coverage persons described in sections 1025(a) or 1026(a) of the Act. Under section 1002(6) of the Act, a “covered person” means (A) any person that engages in offering or providing a consumer financial product or service; and (B) any affiliate of a person described in (A) if such affiliate acts as a service provider to such person.”)
