

Cambridge Analytica and the Meaning of Privacy Harm

Jane Bambauer¹

In 2014 Aleksandr Kogan, a young professor at Cambridge University, made a personality quiz called “This Is Your Digital Life” that attempted to categorize the people who took it into a set of psychological profiles. Almost everyone who took the quiz agreed to let him access not only their public account information –all of it, including every “like”, but also the profiles of their friends, too. That’s how a few hundred thousand survey-takers grew a database of 87 million people.² And even though only those few hundred thousand took the personality quiz that mapped their psychometric profile, the researcher could use their results to build a statistical model that could then predict what the psychometric profiles of the other 86 million Facebook users. The profiles were then provided to the researcher’s company, Cambridge Analytica, which used them as assets to try to get political consulting clients. Like Donald Trump. Facebook changed its policies to restrict how third party companies collect information from Facebook users in 2015, but by the time the change was implemented, Cambridge Analytica had already received the data.

Almost everybody thinks that something went wrong. But defining the harm such that it can elucidate the best course for public policy is an exceedingly difficult task. Some object that Facebook had allowed third party companies to collect users’ data using the consent of their friends, particularly since even those friends who gave consent rarely read or thought through the implications of the disclosures. Some object to the fact that a trivial-seeming personality quiz was used to study and make inferences about the politics and psychology of Facebook users without going through the traditional procedures of academic research. Some object to the fact that third parties can use Facebook users’ data to behaviorally target the distribution of content (especially when the content is politically or ideologically potent).

Each of these concerns deserves careful consideration, but they have divergent implications for how policymakers should respond. Each policy response has potentially broad-sweeping implications for current and future technologies because they address core features of the digital economy. Few people would want to ban *all* of these basic practices, even if they think one or more should be greatly constrained by law. Thus, it is not surprising that even concerned policymakers have splintered and disagreed as the dust from the Cambridge Analytica scandal has settled.

This White Paper uses the Cambridge Analytica episode to explore why the meaning of a privacy-related harm is so difficult to define. The episode reveals that the philosophical fissures

¹ Director, Program on Economics & Privacy, George Mason University Antonin Scalia Law School; Professor of Law, University of Arizona James E. Rogers College of Law. Many thanks to Deven Desai for feedback and guidance on an earlier draft.

² Issie Lapowsky, *Facebook Exposed 87 Million Users to Cambridge Analytica*, WIRED (April 04, 2018), available at <https://www.wired.com/story/facebook-exposed-87-million-users-to-cambridge-analytica/>.

and policy disagreements occur at the foundational layer of conceptualizing privacy. Thus, lawmakers are wise to proceed with caution as they develop new privacy laws.

Theory of Harm 1: Ineffective Consent

The Fair Information Practice Principles, which form the backbone of European Union (EU) privacy law, require notice and affirmative consent before data is collected, shared, or used in a manner that is different from the original purpose of collection.³ A consent requirement propertizes personal information, and treats the person *described* by the data as the owner and controller. Under this model of privacy, a person's loss of control over information that describes them is a cost in itself, and there is no need to search any further for downstream risks or harms.

The dominant narrative in the popular press uses this model of privacy harm. Coverage that characterizes the events as a “leak” or a “breach” treat the data subjects as the presumed rights-holder.⁴ Facebook users who took Kogan's personality quiz did not realize that they were providing access to their *own* Facebook account data, let alone their friends'.

As a matter of description, there is nothing wrong with this account. Abundant evidence shows that people do not read privacy disclosures, and that they have incorrect assumptions about the background law and business of online data collection.⁵ Moreover, even if the survey-takers *did* read and accept the terms of the disclosure, their friends did not.⁶ But as a policy matter, treating lack of consent as a cognizable legal injury in this case would be a major departure from American law, and would have far-reaching implications that conflict with current consumer expectations. The next subsections will explain why this is so.

A. Privacy Interests in “Public” Information

³ Arielle Pades, *What is GDPR and Why Should You Care?* WIRED.COM (May 24, 2018), available at <https://www.wired.com/story/how-gdpr-affects-you/>.


⁴ See, e.g., Cecilia Kang & Sheera Frenkel, *Facebook Says Cambridge Analytica Harvested Data of Up to 87 Million Users*, N.Y. TIMES (April 4, 2018) (originally printed with the headline “Facebook Puts Profile Breach At 87 Million” in the print edition).

⁵ Aleecia M. McDonald & Lorrie Faith Cranor, *The Cost of Reading Privacy Policies*, 41 S. CALIF. L. REV. 543 (2008); Chris Jay Hoofnagle & Jennifer King, *What Californians Understand About Privacy Online*, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1262130; Chris Jay Hoofnagle et al., *How Different are Young Adults from Older Adults When it Comes to Information Privacy Attitudes and Policies?*, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1589864.

⁶ Moreover, even the survey-takers may not have realized that they were providing consent for their friends' data to be harvested for commercial purposes. The context for the personality survey could have been misleading if users believed the data would be collected and used solely for academic research purposes. The survey was presented as an academic research endeavor. However, the explicit terms of service used by Aleksandr Kogan reserved the right to edit, copy, disseminate, publish, transfer, append or merge with other databases, sell, license (by whatever means and on whatever terms) and archive your contribution and data”—terms that violated Facebook's terms of use with app developers. Kevin Poulsen, *Oops! Mark Zuckerberg Surprised to Learn the Terms of Service for ‘Your Digital Life’*, THE DAILY BEAST (April 10, 2018), available at <https://www.thedailybeast.com/oops-mark-zuckerberg-surprised-to-learn-the-terms-of-service-for-your-digital-life>.

First, the data collected by Cambridge Analytica was more limited than the popular media suggests—particularly for friends of the survey-takers. Facebook’s application program interface (“API”) allowed third party app developers to collect the app user’s friends’ information if it was part of their public profile⁷ or was already set to be “shared with everyone,” to use the Facebook parlance. (This includes page likes, which are set to public as a default, but does not include likes on individual posts.) The only information collected about friends that was not already available on the public world-wide web was the birthdate and current city of the friends. The data was therefore much less rich than some reports have suggested.⁸

For example, my information was included in the Cambridge Analytica database because somebody I am friends with took the personality quiz. Facebook has acknowledged the disclosure of my public profile:

 What banned apps may have had access to my info?			
Based on our ongoing investigation, it appears that the following apps banned by Facebook for possible data misuse may have had access to your info.			
App Name	Why were they removed?	How was my info accessed?	What info could the app access?
This Is Your Digital Life Removed: 12/17/2015	The app was removed for selling information to a third party.	A friend may have shared your info as part of logging into the app.	Your public profile , Page likes, birthday and current city

Thus, the data that Cambridge Analytica collected was limited in scope, and usually already publically accessible. In theory the same information could have been scraped from the public web without any cooperation from Facebook.

This does not end the privacy discussion, of course. Most people agree that there are degrees of obscurity⁹, and that old rules that rely on a strong dichotomy between public and private may not be a good fit for modern information management. Nevertheless, commentators have tended to focus on the large number of people whose data wound up in the hands of Cambridge Analytica without recognizing that the information collected was already accessible to third parties.

B. *Conflicting Consumer Interests*

A property right to control personal information would certainly limit the activities of Facebook, Cambridge Analytica, and other intermediaries. But it would also limit the choices for other

⁷ Facebook requires all users to make some basic information public, such as name, age range, language and country.

⁸ See, e.g., Carole Cadwalladr & Emma Graham-Harrison, *How Cambridge Analytica Turned Facebook ‘Likes’ Into a Lucrative Political Tool*, THE GUARDIAN (Mar 17, 2018).

⁹ This dimension of privacy was popularized by Woody Hartzog. See Woodrow Hartzog & Frederic D. Stutzman, *Obscurity by Design*, 88 WASH. L. REV. 385 (2013).

consumers. For example, existing biometric privacy laws in Illinois prevent consumers from being able to purchase home security systems that use facial recognition to identify visitors because the visitors, rather than the homeowner, have exclusive control over their biometric data.¹⁰ And the same law has exposed Facebook to billions of dollars in liability for its innocuous automatic photo-tag suggestion tool.¹¹ New general data privacy laws could create similar problems, and would require exemptions to avoid creating liability for individuals who disclose personal details about others on social media or even to journalists. These trade-offs might turn out to be the right ones, but it will be important for regulators to recognize that privacy laws have ambiguous effects. They are not uniformly pro-consumer.

C. Effective Notice

Whether a Facebook user has the right to provide his friends' information or not, consent to provide that information requires effective notice about its collection. The notice that Facebook provided to third party app users were flawed at the time Kogan was collecting the Cambridge Analytica data. While the notice made clear that the app requests (or requires) access to the user's public profile information, the data collected on friends was listed under the heading "Access my basic information," and the request was made for "lists of friends" without reference to those friends' public profile information.



Many commentators have focused on the potential for deception and confusion because of the wording of these requests.¹² I tend to agree that this wording leaves much to be desired if the goal is to educate individual end users, though it is at least plausible that "lists of friends" is sufficient to describe access to the profile information of friends that is already exposed to everybody on the world wide web.

¹⁰ Ally Marotti, *Proposed Changes to Illinois' Biometric Law Concern Privacy Advocates*, CHI. TRIBUNE (April 10, 2018), available at <https://www.chicagotribune.com/business/ct-biz-illinois-biometrics-bills-20180409-story.html>.

¹¹ *Patel v. Facebook*, 290 F.Supp.3d 948 (2018).

¹² Elizabeth Dwoskin & Tony Romm, *Facebook's Rules for Accessing User Data Lured More than Just Cambridge Analytica*, WASHINGTON POST (March 19, 2018), available at https://www.washingtonpost.com/business/economy/facebooks-rules-for-accessing-user-data-lured-more-than-just-cambridge-analytica/2018/03/19/31f6979c-658e-43d6-a71f-afdd8bf1308b_story.html?utm_term=.f9c3c09d2c31.

Assuming that this wording is inadequate to put users on meaningful notice, Facebook broke its own promises and assurances that it would never disclose a user's data without consent. A company that violates its own privacy policies has deceived its users, and possibly lured them away from competitors who have better privacy practices. If the Federal Trade Commission (FTC) believes that Facebook's API disclosures were misleading, the Commission could easily bring an enforcement action based on deception.¹³

But the language of a privacy disclosure has little relevance to foundational theories of privacy law. The drafting of privacy notices has almost no effect on behavior.¹⁴ Consumers rarely alter the defaults in privacy settings¹⁵, and even though a greater proportion of users are starting to work with privacy settings¹⁶, they rarely forego a service or benefit that they would otherwise want in order to protect their privacy.¹⁷ Thus, whatever the wording of a privacy notice, the most important factor for consumer data collection is whether third party companies can collect data as a default or not.

D. Would You Like Your Services With or Without Friction?

The determinative power of defaults is consistent with two conflicting explanations. The first is that despite our collective protestations, revealed preferences show that Americans do not value privacy greatly. They do not value it enough to pay even small amounts of money, or to forego even trivial benefits, to preserve it.¹⁸ The other, opposite theory is that Americans care deeply about privacy, but they lack information about the costs and benefits of data transactions and about how much information has already been collected or inferred about them. Without insight about which services are worth the loss of control and which are not, consumers are resigned in their privacy vigilance.¹⁹

Proponents of the revealed preferences theory prefer to set defaults to allow for frictionless data transfers, while proponents of the lost autonomy theory usually prefer to set defaults the other way—to protect privacy unless a consumer opts into data collection. Given the power of these defaults, policymakers who adopt a property model for personal data have only two options. If data transfers can occur as a default, data will be very fluid and quickly find its way into the

¹³ Or one based on Facebook's existing FTC settlement, for that matter. *Id.*

¹⁴ Lior J. Strahilevitz & Matthew B. Kugler, *Is Privacy Policy Language Irrelevant to Consumers?*, 45 J. LEG. STUD. 569 (2016); Adam S. Chilton & Omri Ben-Shahar, *Simplification of Privacy Disclosures: An Experimental Test*, 45 J. LEG. STUD. 41 (2016).

¹⁵ Adam S. Chilton & Omri Ben-Shahar, *Simplification of Privacy Disclosures: An Experimental Test*, 45 J. LEG. STUD. 41 (2016); Eric J. Johnson et al., *Defaults, Framing and Privacy: Why Opting In-Opting Out*, 13 MARKETING LETTERS 5 (2002).

¹⁶ Andrew Perrin, *Americans Are Changing Their Relationship with Facebook*, PEW RESEARCH CENTER, available at <http://www.pewresearch.org/fact-tank/2018/09/05/americans-are-changing-their-relationship-with-facebook/>.

¹⁷ Lior J. Strahilevitz & Matthew B. Kugler, *Is Privacy Policy Language Irrelevant to Consumers?*, 45 J. LEG. STUD. 569 (2016).

¹⁸ *Id.*; Alessandro Acquisti, *Privacy and Human Behavior in the Age of Information*, 347 SCIENCE 509 (2015) (providing a summary of related scholarship.)

¹⁹ JOSEPH TUROW ET AL., *THE TRADEOFF FALLACY* (2015), available at https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf.

servers of companies. If the default is set to privacy, personal data will become very sticky forms of property, baking in an assumption that consumers are better off if companies cannot access their data.

Both of these options are suboptimal. High friction privacy is likely to do a disservice to consumers by reducing innovation and competition. For example, a study of privacy defaults in the laws regulating home loans in the San Francisco Bay area found that loan applicants living in the counties that set privacy as the default paid higher interest rates and defaulted at greater rates than the counties that set data flow as the default, even after controlling for confounders, because banks could not match applicants to loans as well and naturally passed the costs of risk along to the consumer.²⁰ But low friction privacy is unlikely to thwart genuine problems for the consumer where they do occur.

The fact that most people do not change defaults even when it is fairly easy to do is a clue that the property model is not a good fit for privacy. The standard justification for property rights assumes that a property interest creates the right incentives to make sure that an asset is owned by the actor who can make the best, most productive use of it. It allows each actor's private information to guide their decisions. If consumers don't have special insight—if they do not know the potential risks and benefits that sharing data will have to them—the foundational assumptions of a property model are wrong.

This helps explain why American policymakers have resisted the international trend to adopt something like the EU's General Data Protection Regulation (GDPR). At the same time, few people would disagree with the assumption that privacy has value—that there is a rational and very human need for respite from judgment and from unintended consequences in some contexts. Thus, some policymakers including the FTC have begun to focus less on notice and consent, and more on defining downstream risks and harms.

The next sections explore some theories of informational injuries stemming from the Cambridge Analytica scandal that could be corrected by regulators without expecting data subjects to protect themselves.

Theory of Harm 2: Surreptitious Inference

Cambridge Analytica provoked distrust through its use – not just its collection—of the Facebook data. Facebook users who took the My Digital Life survey unwittingly helped improve Kogan's psychometric profiling model. And then, the improved model was applied to their friends' data to predict the friends' psychometric profiles. Is studying people and predicting their unobserved qualities without informed consent a harm that policymakers should address?

Some may recall a different Facebook controversy that involved surreptitious research a few years ago when an academic journal article revealed that Facebook had worked with outside

²⁰ Jin-Hyuk Kim & Liad Wagman, *Screening Incentives and Privacy Protection in Financial Markets: A Theoretical and Empirical Analysis*, 46 RAND J. OF ECON. 1 (2015). This is consistent with the more general phenomenon of risk-based lending markets. See Wendy Edelberg, *Risk-Based Pricing of Interest Rates for Consumer Loans*, 53 J. MONETARY ECON. 2283 (2006).

researchers to secretly test whether emotions are “contagious” through the newsfeed algorithm.²¹ Facebook set up a randomized controlled trial (“RCT”) on a subset of its users to investigate whether changes in the newsfeed algorithm to suppress either happy or unhappy messages wound up changing the posting behavior of the users. The results—that there was a very small effect—were less dramatic than the public relations fallout. Many Facebook users resented being treated like lab rats.²²

Cambridge Analytica did not run experiments on people, but they did use personal information to study latent political and psychological qualities of people without their permission. Perhaps policymakers should reduce the opportunity for firms to engage in secret research and inference by banning it or mandating something like the IRB procedures that academic research institutions have to use.²³

If research and inference performed without the subject’s knowledge is the basis for harm, the obvious solution is to require notice and consent. (Theory 2 therefore shares some similarities to the property model described in Theory 1, but applying the consent requirement to *uses* of personal data rather than to its collection and dissemination.) Because a significant set of the general public have reacted negatively when surreptitious research has been uncovered, restricting research might seem like a narrow and wise course for policymakers. But in fact, legal restrictions on research would be neither narrow nor wise.

Inference and experimentation are at the heart of everything that we love about the Internet and smart devices. Websites and Internet service providers use randomized controlled experiments so often that the industry even has its own cute name for them- “A/B Testing.” By applying inferences quietly and automatically, firms swiftly improve the accuracy and efficiency of their services, usually by finding correlations or repurposing data that end users would never have imagined. Laws that ban or heavily regulate inference would end the digital revolution as we know it. And if this isn’t reason enough to abandon this theory of harm, there is also the problem that a public law regulating research and inference is very likely to conflict with the First Amendment.²⁴

Cambridge Analytica is a poor test for public sentiment related to research. The public does not trust that the motivation for Cambridge Analytica’s psychometric inference were consistent with consumer welfare. Assuming that there was a conflict between the interests of Cambridge Analytica and the people it studied, their research may be unrepresentative, given the ubiquity of

²¹ Adam D.I. Kramer et al., Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks, 111 Proc. Nat’l Acad. Sci. 8788, 8788 (2014).

²² Dino Grandoni, *You May Have Been a Lab Rat in a Huge Facebook Experiment*, HUFFINGTONPOST.COM (June 29, 2014), available at https://www.huffingtonpost.com/2014/06/29/facebook-experiment-psychological_n_5540018.html.

²³ For a thoughtful argument along these lines, see James Grimmelman, *Law and Ethics of Experiments on Social Media Users*, 13 COLO. TECH. L. J. 219 (2015); Raquel Benbunan-Fich, *The Ethics of Online Research with Unsuspecting Users: From A/B Testing to C/D Experimentation*, 13 RESEARCH ETHICS 200 (2017).

²⁴ Indeed, Philip Hamburger has argued that even the Institutional Review Boards that oversee research at federally funded academic institutions may violate the First Amendment. Philip Hamburger, *The New Censorship: Institutional Review Boards*, 6 SUP. CT. REV. 271 (2004). See also Jane Bambauer, *Is Data Speech?*, 66 STAN. L. REV. 57 (2014).

inference in the modern economy. A research insight into the psychology of a person or any other latent quality can be exploited for a range of purposes—some beneficial, some harmless, some presumably malignant.

Given the historical successes and great potential for the beneficial uses of inference, it is incumbent on policymakers to regulate not research itself, but harmful applications of research. This, of course, shifts the definition of information harm further downstream, requiring regulators to more specifically define a use of data that puts its subjects in peril.

I will discuss two possible elucidations of Cambridge Analytica's misuse of inference in the next section. But before moving on, it is also worth acknowledging that the power of Cambridge Analytica's psychometric profiling has been exaggerated. Neither Cambridge Analytica's algorithm or anybody else's do a good job categorizing people into political-psychological profiles. Kogan estimated a 30% success rate at sorting Facebook users into a small number of crude categories, and this is in line with other attempts to do so.²⁵ His model, like others in the business of political prediction, relies most heavily on basic demographic information, suggesting that the Facebook data added very little predictive value. Cambridge Analytica may have been selling political Snake Oil to the Trump campaign.

Anybody who fears that Machine Learning will know the future should take heart. One of the enduring truths in sociology and behavioral research is that even with lots and lots of historical data, people are hard to predict. So while an algorithm may indeed predict our future behavior better than we could predict it ourselves, this is only because we are so bad at predicting ourselves. All indicators suggest that humans are still quite immune to reductivism.

Theory of Harm 3: Behavioral Targeting by Third Parties

Traveling further downstream in the information flow, the next natural landing place for exploration is in the use of Big Data revelations. If the study of consumers is not a harm in itself, the insights that come from data analytics could be misused for manipulative and nefarious purposes. Cambridge Analytica, for example, may have been able to exploit their inferences about Facebook users' psychological profile in order to craft advertisements to stoke their psychological fears, insecurities, and vulnerabilities.

To address this possibility, policymakers could consider restricting the ability to use Facebook user data to hyper-customize the content that is served to them. This restriction could apply to all third parties or to some subset of them. Let's consider each of these options.

A. Restricting Third Party Targeting

Facebook facilitates two types of behavioral targeting. It's primary function and value for consumers is to not only gather the content of each user's social network, but then to curate the

²⁵ Matthew Hindman, *How Cambridge Analytica's Facebook Targeting Model Really Worked—According to the Person Who Built It*, THECONVERSATION.COM (March 30, 2018), available at <https://theconversation.com/how-cambridge-analyticas-facebook-targeting-model-really-worked-according-to-the-person-who-built-it-94078>. See also Evan Halper, *Was Cambridge Analytica a Digital Svengali or Snake-Oil Salesman?*, L.A. TIMES (Mar 21, 2018).

content based on that user's previous browsing behavior and revealed preferences. This service has been scrutinized and criticized by many commentators (perhaps most famously in Eli Pariser's *The Filter Bubble*²⁶), but whatever pathologies might exist in this service are media problems rather than privacy problems. The management of users' social networks is the *raison d'être* of Facebook.

But Facebook also facilitates another type of behavioral targeting—that by third parties who are not part of a user's social network (advertisers and other content producers). Cambridge Analytica used this route to access and influence Facebook users. Thus, during the congressional hearings following the breaking of the scandal, some members of Congress latched onto the idea that Facebook could be persuaded or forced to stop selling behavioral targeting services. Mark Zuckerberg was asked about changing the business model to a fee-based subscription service for Facebook users who want to avoid behavioral targeting by advertisers. Zuckerberg rejected the suggestion.²⁷

Why did Zuckerberg reject an idea that had so much political and popular support? It could be that he is tone-deaf or insensitive, but a more likely explanation is that Zuckerberg has inside information about the privacy paradox—and more specifically, about the gap between consumers' stated preferences for privacy and their actual willingness to pay for it. Advertising that is targeted to specific users based on their personal data is significantly more valuable than advertising that is not. It is much more likely to receive clicks and “conversions”—that is to say, purchases.²⁸ Thus, Facebook, Google, and other content intermediaries can command higher prices for ad space when it can be targeted based on the personal characteristics and histories of its users. While consumers value their privacy, they would not be willing to make up the difference in the revenues that would be lost from banning behavioral targeting.

Facebook, Google, and other tech giants who offer services to consumers for no monetary charge are built on a business model of tailored advertising. While Congress is under great pressure from voters to fix Facebook, it is not yet ready to kill it.

B. Restricting Third-Party Targeting Based on Content or Identity

²⁶ ELI PARISER, *THE FILTER BUBBLE* (2011).

²⁷ Michele Castillo, *Mark Zuckerberg Hints That Facebook Has Considered a Paid Version*, CNBC.COM (April 10, 2018) ("To be clear, we don't offer an option today for people to pay to not show ads. We think offering people an ad-supported service is the most aligned with our mission of trying to connect everyone in the world, because we want to offer a free service that everyone can afford. That's the only way we can reach billions of people.").

²⁸ The “lift” on conversion rates from Facebook's behavioral advertising is about 73% on average. Brett R. Gordon et al., *A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook*, *MARKETING SCIENCE*, *31 (forthcoming), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3033144. Research on advertising without behavioral targeting has found an 8% lift for Google Display Network ads. See Garrett A. Johnson et al., *The Online Display Ad Effectiveness Funnel & Carryover: Lessons from 432 Field Experiments*, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2701578.

Even if Facebook keeps its advertising-driven business model, perhaps it should be prohibited from giving access to user data or in-house targeting services to advertisers who are pursuing an overtly political or ideological agenda. This sort of ban would excise the practices that have received the most blame for contaminating American politics, and would be better tailored to the anxiety that is motivating much of the consternation—the outcome of the 2016 presidential election.

This approach has a number of challenges, too, not the least of which is that it reverses the standard hierarchy of free speech theory (which usually puts political speech at the center of Constitutional protection.) But even overlooking First Amendment problems, implementation would be hazardous. Differentiating ideological from apolitical content is nearly impossible when everything from global warming to football games have been politicized. Banning the targeted delivery of fake news clickbait also requires difficult distinctions between reasonable speculation and falsehoods, or between persuasion and manipulation.

Another problem with the theory of harm based on political distortion is that it throws policymakers into a course of treatment before a diagnosis has been worked out. The Cambridge Analytica scandal galvanized scholars and advocacy groups who already believed Facebook was unwittingly complicit in a Fake News catastrophe, contributing to the miseducation of the American public. But quantitative studies of media exposure suggest that while a sizeable proportion of the public (about a quarter) saw at least one fake news story during the weeks leading up to the 2016 elections, fake news websites constituted only a small fraction of news consumption—about 2.6%. Only 10% of American voters had intensive, repeated exposure to fake news sites like Breitbart, and this group had preexisting beliefs at the extreme ends of the ideological spectrum (mostly on the conservative side).²⁹ And although readers exposed to fake news content were more likely than not to believe the false facts that were reported, the tendency to believe fake news was highly correlated with the content already matching the reader's political beliefs and with the reader having an ideologically segregated social circle.³⁰

With these descriptive statistics, it is clear that it will be very difficult for researchers to know confidently whether the distribution of propaganda and fake news causes ideological polarization or whether the causation is reversed—that an already polarized public is more drawn to content that is consistent with their world views. Indeed, one of the tactics of Russian interference involved simply tapping into and amplifying genuine issues of political disagreement.³¹ This is not to say that Facebook had no marginal effect on voting behavior, and Facebook's self-motivated efforts to weed out Russian-generated or demonstrably false news content is commendable. But today it is far from clear that Facebook content changed political attitudes or turnout enough to affect the outcome of the 2016 election.

²⁹ Andrew Guess et al., *Selective Exposure to Misinformation: Evidence from the Consumption of Fake News During the 2016 U.S. Presidential Campaign*, available at <http://www.ask-force.org/web/Fundamentalists/Guess-Selective-Exposure-to-Misinformation-Evidence-Presidential-Campaign-2018.pdf>.

³⁰ Hunt Allcott & Matthew Gentzkow, *Social Media and Fake News in the 2016 Election*, 31 J. ECON. PERSPECTIVES 211 (2017).

³¹ Kate Conger & Charlie Savage, *How Fake Influence Campaigns on Facebook Lured Real People*, N.Y. TIMES (Aug 2, 2018).

Assuming that the spread of fake news on Facebook did have an appreciable effect on voting behavior, a government intervention has its own risks. The conspiracy theorists among us probably do not respond well to government-initiated censorship. Indeed, Trump supporters would be justified in suspecting that the strong reaction to Cambridge Analytica's practices was partly motivated by politics, as the first presidential campaign to make aggressive use of Facebook's API for snowball data collection was not Trump's but Obama's. Supporters who signed up for Obama's campaign mailing list were urged to download the Obama campaign's Facebook app which collected not only their Facebook data, but that of their friends as well.³²

Conclusion

This white paper has laid out some avenues for defining privacy-related harms in the wake of Cambridge Analytica, and has identified significant (though not necessarily insurmountable) problems with each. Notice and consent models that treat data as the property of people who are described by it tend to bog down innovation with transaction costs and also are unlikely to prevent harm to consumers who have limited information about downstream uses anyway. Specific theories of downstream risks are also hard to nail down. The surreptitious study of consumers is "creepy," but is also an essential component of the Fourth Industrial Revolution (not to mention free inquiry.) And attempts to limit how Big Data analytics are used will require more spadework to separate acceptable persuasion from unacceptable manipulation.

The policy discussion in the wake of Cambridge Analytica is best understood in context of two other phenomena. First, the scandal broke during efforts to understand the outcome of a national election that was surprising and profoundly disturbing to many people (this author included.) Thus, the details of the data collection and use practices have inevitably been viewed through a particular lens—one clouded by a desperate desire to find and fix problems in the democratic process. Second, when societies are under stress, people habitually see existential threats from

³²Kalev Leetaru, *Why Are We only Now Talking About Facebook and Elections?*, FORBES.COM (Mar 19, 2018), available at <https://www.forbes.com/sites/kalevleetaru/2018/03/19/why-are-we-only-now-talking-about-facebook-and-elections/#6118f1914838>; Mark Sullivan, *Obama Campaign's 'Targeted Share' App Also Used Facebook Data from Millions of Unknowing Users*, FAST COMPANY (Mar 20, 2018), available at <https://www.fastcompany.com/40546816/obama-campaigns-targeted-share-app-also-used-facebook-data-from-millions-of-unknowing-users>. Attempts to distinguish the Obama campaign's practices seem rather thin. The chief scientist for President Obama's 2012 campaign told reporters that the practices were different from Cambridge Analytica's because the Obama campaign did not directly contact friends of the app users. "All we could do was ask our 'primary' supporters to contact their friends and we would recommend who those friends were based on the data they allowed us to access." Ken Thomas, *Obama Campaign Advisers Say They Used Facebook Data Properly*, CHI. TRIB. (Mar 22, 2018), available at <https://www.chicagotribune.com/news/nationworld/politics/ct-obama-campaign-facebook-data-20180322-story.html>. But the similarities to Cambridge Analytica are arguably more important than the differences. The Obama campaign relied on the consent of their primary contacts to access data of their friends, and then analyzed that data to find new potential supporters. The Obama campaign also used microtargeting of ad delivery on cable television, and while they didn't use Facebook's ad placement, this is in part because it would be *more* effective for a targeted message to come directly from a real friend rather than through Facebook. "[T]he campaign's ultimate goal was to *deputize* the closest Obama-supporting friends of voters who were wavering in their affections for the president." Jim Rutenberg, *Data You Can Believe In*, N.Y. TIMES (June 20, 2013) (emphasis added), available at <https://www.nytimes.com/2013/06/23/magazine/the-obama-campaigns-digital-masterminds-cash-in.html>.

two sources: immigrants and technology.³³ Level-headed commenters and public intellectuals see clearly the effects of xenophobia on Trump-era public policy, but technophobia is also likely to affect political debate. Thus, broad-sweeping policy should await a more stable theory of privacy harm and an empirical base of relevant evidence.

³³ BRYAN CAPLAN, *THE MYTH OF THE RATIONAL VOTER* (2007).