

HEARING BEFORE THE UNITED STATES SENATE
COMMITTEE ON THE JUDICIARY

Written Testimony of Eitan Hersh

May 16, 2018

Introduction

Chairman Grassley, Ranking Member Feinstein, and distinguished members of the committee, it is an honor to offer testimony before the Judiciary Committee. My name is Eitan Hersh. I am a professor of political science at Tufts University. My research and teaching focus on civic participation and the relationship between laws, political strategies, and the behavior of voters. Much of my research utilizes large databases of information about voters such as public voter registration records. I have used these databases as an expert in litigation involving voting rights and as a scholar of political behavior.

In 2015, I published a book called *Hacking the Electorate: How Campaigns Perceive Voters*, which takes an in-depth look at the databases used in the 2008-2014 election campaigns. I describe the data campaigns were using in those elections, how they accessed the data, and the effects of the data practices on political outcomes. Much of the data in the study comes from Democratic organizations, including from Catalist, the 2012 Obama re-election campaign, NGP-VAN, as well as from interviews with political data staffers from both parties. Some of the data comes from experiments I conducted on the effectiveness of campaign ads.

Summary of Testimony

The recent controversy over Cambridge Analytica and its use of Facebook data in support of the Donald Trump presidential campaign has raised a number of serious concerns for the American public. These concerns include foreign interference in US elections, the personal privacy of Facebook users, third-party misuse of Facebook data, and voter targeting practices. In considering the controversy over Cambridge Analytica, I endeavor to provide testimony about one of these issues: voter targeting.

First, I will describe voter targeting practices. Based on the information I have seen from public reports about Cambridge Analytica, it is my opinion that its targeting practices in 2016 ought not to be a major cause for concern in terms of unduly influencing the election outcome. Second, I will explain the gaps in our knowledge about the effects of social media-based targeting. Much more could be learned by impartial researchers to determine the power of targeting tools used in the 2016 election and, more importantly, the landscape of targeting in the coming years. In order

for researchers to learn these things, they will need access to data held by Facebook. Third, I will suggest that those interested in the effect of social media platforms on electoral politics should focus not only on the *supply* of provocative political information from campaigns and firms like Cambridge Analytica, but also on the *demand* for provocative information from American citizens.

The Limits of Targeting in Political Campaigns

In every election, the news media exaggerate the technological feats of political campaigns. Back in 2004, the *New York Times* reported that parties “could divine my likely views on taxes, law enforcement, abortion, and global warming.” In 2008, the *Washington Post* reported that campaigns “can figure out what John Smith at 286 Main Street is thinking.” In 2012, a *CNN* headline read, “Microtargeting: How campaigns know you better than you know yourself.”¹ The actual capability of campaigns in these years never lived up to the hype. In hindsight, this is obvious. At the time, it would not have been obvious to anyone who wasn’t well-versed in the campaigns’ actual data practices. In truth, the most important things campaigns know about voters are contained in public records, such as party affiliation as recorded in voter registration files.

The news media are prone to overstating the power and sophistication of campaign techniques, for at least three reasons. First, news readers are more interested in learning about the promise of technology than about its limitations. Second, after an election, there is always a demand to figure out why the winning campaign won. The latest technology used by the winning campaign is often a good storyline, even if it’s false. Finally, campaign consultants have a business interest in appearing to offer a special product to future clients, and so they are often eager to embellish their role in quotes to the media.

The technological landscape has changed since 2012. As technology changes, there is a legitimate worry that the latest innovation really is different from anything that has been done in the past. In the present context, there is a collective anxiety over whether Cambridge Analytica used Facebook data to construct targeting models that cross the line from persuasion into manipulation: are the strategies employed more akin to a traditional ad offering reasons to support a candidate or are they more akin to a subliminal message intended to deceive?

¹ Jon Gertner, “The Very, Very Personal is the Political,” *New York Times Magazine*, February 15, 2004; Steven Levy, “In Every Voter, A ‘Microtarget,’” *Washington Post*, April 23, 2008; Allison Brennan, “How Campaigns Know You Better Than You Know Yourself,” *CNN*, November 5, 2012.

Campaigns generally endeavor to mobilize and to persuade potential voters. Mobilization entails finding likely supporters who might not vote without encouragement and encouraging them to vote. Advancements in data and methods have improved the ability for campaigns to mobilize. Information contained in public records, such as party affiliation, race, gender, age, and geography is very informative in identifying which voters will support Democrats and Republicans. Prior vote history, which is available as a public record in all states, as well as information such as likely marital status, age, and electoral context, give campaigns a good sense of which voters might not show up to vote without a reminder from a campaign. Databases containing personal information of this kind, as well as new strategies developed through experimentation, have improved the capabilities of campaigns to mobilize.

Commercial data – such as information about purchasing habits or leisure interests – can also help campaigns with mobilization, but their use in the past has been limited. In *Hacking the Electorate*, I found that commercial data did not turn out to be very useful to campaigns. Even while campaigns touted the hundreds or thousands of data points they had on individuals, campaigns’ predictive models did not rely very much on these fields. Relative to information like age, gender, race, and party affiliation, commercial measures of product preferences did not add very much explanatory power about Americans’ voting behavior.

Some of the reasons why commercial data wasn’t useful in the past no longer generally applies to online data supplied by firms like Google and Facebook. For instance, offline commercial data is often inaccurate, out-of-date, and not available for a large share of the population. Facebook data doesn’t have these constraints.² In *Hacking the Electorate*, I wrote,

Facebook allows advertisers to target user accounts based on profile segments, but the company does not allow clients to extract lists of users identified by name and by traits like self-reported ideology. If they did, this data field would be of substantial use to campaigns, allowing them to perceive the self-reported ideological disposition of millions of Americans. Campaigns could then engage voters in direct contacting strategies based on these perceptions. This is the kind of commercial data field that exists but has so far not been shared with or sold to campaigns.

Since I wrote that in 2015, it is apparent that Facebook data has been extracted and used to inform targeting models. This could improve mobilization efforts. For instance, many voters live in states that do not record party affiliation on voter registration databases. In these states, it is often hard for campaigns to figure out which voters are Democrats and which are Republicans.

² While it may be broadly true that Facebook data has wide coverage and accurate records about the American public, it is worth noting that the demographic group credited for President Trump’s victory, the white working class, may be the least likely group to use Facebook and therefore the least likely to afford campaigns with a digital path to voter engagement. In other words, Facebook data may reveal less about the white working class than about any other segment of the American public. According to a recent report from Pew, voters who are older, who are white, who have no college education, who are rural, and who are male are less likely to use Facebook than their demographic counterparts. See Pew Research Center for Internet and Technology, Social Media Fact Sheet, February 5, 2018, <http://www.pewinternet.org/fact-sheet/social-media/>.

Facebook data – for instance a self-reported measure of ideological liberalism or conservatism (a field Facebook collects from users) – would improve a campaign’s ability to identify its supporters.

At the same time, some of the reasons why commercial data wasn’t useful in 2008 or 2012 still render commercial data of limited use today. Many commercial fields simply are not highly correlated with political dispositions. And even those that are might not provide *added* information to a campaign’s predictive models. As an example, consider boat ownership. Boat ownership, according to data I studied, is correlated with being a Republican. However, it provides little help to a campaign because once a campaign knows a person is, for example, a 55-year-old white man living in a wealthy Republican-leaning seaside enclave, the campaign already predicts this person is a Republican. The commercial field tells the campaign nothing new about the voter’s partisan affinity. Because campaigns have rich data on demographics and neighborhoods stemming from public records, and because some of the biggest cleavages in American politics fall on simple demographic lines of age, race, gender, and geography, commercial data doesn’t always add as much value to campaign mobilization efforts as would appear at first blush.

Quite different from a campaign’s efforts to mobilize (or demobilize) is the strategy of persuasion. Persuasion is a campaign’s attempt to find citizens who are likely to vote but are either uncertain who they will vote for or are planning to support the other side, and to convey messages to change these voters’ minds.

Persuasion is different from mobilization in that it is much more difficult. Persuadability is an unstable disposition. Whereas a person who supported Republicans last cycle is likely to support Republicans this cycle, a person who was persuadable yesterday might not be persuadable today. There is no one subset of the electorate that is all the time susceptible to persuasion. Depending on the exact message, messenger, and context, a person may be persuadable or not persuadable. This makes it difficult for campaigns and political parties to learn, election to election, or even day to day, about how to persuade voters. What worked last time may not work this time.

Moreover, a persuasion effect is quick to decay. A message seen by a voter may be persuasive for a fleeting moment and then is lost in the cacophony of political ads, news, and posts that fill a Facebook feed. Finally, campaigns learn a lot about the effectiveness of mobilization techniques because whether someone ended up voting or not is a public record. For persuasion, campaigns do not typically learn which candidate the voters ended up supporting, on account of the secret ballot. This again makes it hard for parties, campaigns, and consultants to build up a set of best practices for persuasion.

A large part of the story surrounding Cambridge Analytica is its efforts at persuasion. In a 2016 video presentation, Alexander Nix (former CEO of Cambridge Analytica) described his firm’s work on persuasion on behalf of Senator Ted Cruz’s presidential primary campaign.³ Mr. Nix

³ Alexander Nix “Cambridge Analytica – The Power of Big Data and Psychographics” Concordia Summit Youtube, September 27, 2016, <https://www.youtube.com/watch?v=n8Dd5aVXLcC>

described a persuasion strategy as follows. First, he defines a persuasion universe as people who will vote but the campaign isn't sure who they will vote for. Second, he notes that Cambridge Analytica's psychographic models tell him that these voters are "very low in neuroticism, quite low in openness, and slightly conscientious." Third, he subsets further to a group of individuals who are predicted to care about gun rights. And he concludes, "Now we know that we need a message on gun rights, it needs to be a persuasion message, and it needs to be nuanced according to the certain personality that we're interested in."

Nearly everything Mr. Nix articulates here is not new. Based on what we know from past work, it is also likely to have been ineffective. Cambridge Analytica's definition of a persuadable voter is someone who is likely to vote but the campaign isn't sure who they will vote for. This is a common campaign convention for defining persuadability. It also bears virtually no relationship to which voters are actually persuadable, undecided, or cross-pressured on issues, as I discuss in *Hacking the Electorate*.

When doing persuasion, a campaign is trying to identify individuals who are responsive to a message. That is, a persuadable voter is someone whose opinions will change based on hearing or reading new information. Simply defining a target list as people who are likely to vote but who the campaign doesn't know for whom they will vote is an extraordinarily rough proxy for persuadable voters. It is a proxy that many campaigns have long used for the simple reason that the psychological disposition of persuadability in a given moment for a given candidate is hard to measure. Cambridge Analytica's strategy of contacting likely voters who are not surely supportive of one candidate over the other but who support gun rights and who are predicted to bear a particular personality trait is likely to give them very little traction in moving voters' opinions. And indeed, I have seen no evidence presented by the firm or by anyone suggesting the firm's strategies were effective at doing this.⁴

The new component of targeting described by Mr. Nix and discussed at length in the media is psychological profiling. Apparently, Cambridge Analytica obtained Facebook data and used it in combination with survey responses to predict personality traits like neuroticism and openness. It could use predictions of these traits to target voters.

As many journalists have observed,⁵ building a psychological profile by connecting Facebook "likes" to survey respondents who took a personality test would lead to inaccurate predictions.

⁴An important note: the fact that Donald Trump won the presidency after contracting with Cambridge Analytica or that Senator Cruz increased his name recognition in Iowa following his relationship with Cambridge Analytica do not count as evidence that Cambridge Analytica was effective. Many other things were happening during these campaigns. Cambridge Analytica or Facebook could plausibly demonstrate effectiveness of the targeting strategy through use of experimental techniques. As far as I know, neither firm has publicly reported experimental evidence.

⁵ E.g., Antonio Garcia Martinez, "The Noisy Fallacies of Psychographic Targeting," *Wired*, March 19, 2018; Brian Resnick, "Cambridge Analytica's 'psychographic microtargeting,': what's bullsh*t and what's legit," *Vox*, March 26, 2018.

Facebook “likes” might be correlated with traits like openness and neuroticism, but the correlation is likely to be weak. The weak correlation means that the prediction will have lots of false positives – namely, people who Cambridge Analytica predicts will have a trait but who actually don’t have that trait.

To put some numbers on this, we can use models of racial identity as a baseline. Racial identity seems like it ought to be a relatively easy trait to predict because it is stable and because available information such as a person’s name and where he or she lives is correlated with his or her racial identity. In campaign targeting models I have studied, predictions of which voters are black or Hispanic are wrong about 25-30% of the time. Models of traits such as issue positions or personality traits are likely to be much less accurate. They are less accurate because they are less stable and because available information like demographic correlates and Facebook “likes” are probably only weakly related to them.

The problem for campaigns with these messy predictions is that voters may not like being “mis-targeted” by receiving a targeted ad not designed for them. In a series of experiments, a colleague and I found that voters penalize candidates for mis-targeting such that any gains made through a successful target are often canceled out by losses attributable to mistargets.⁶ For instance, messages intended for a religious group, a racial group, or an issue group like gun owners do not go over very well when they are presented to people who are not in those segments of the population.

In summary, given research on the difficulty in persuading voters, the difficulty in accurately pinpointing nuanced traits like personality types, and the information revealed to the public about what Cambridge Analytica did, I am skeptical that Cambridge Analytica manipulated voters in a way that affected the election.

Independent Researchers Need More Data

The skepticism I offer comes with a high degree of uncertainty. There’s a lot the public doesn’t know. We don’t know what strategies were actually employed beyond what was described publicly. And we don’t know with certainty how effective these strategies were. We don’t know the full extent of Facebook data used by third parties. We also don’t collectively understand where the line is between targeting that is tolerable and targeting that crosses the line into subconscious manipulation that requires a policy response. Behind the Cambridge Analytica/Facebook scandal is an understandable anxiety about where that line is and how we would even know if it has been crossed.

⁶ Eitan Hersh and Brian Schaffner, “Targeted Campaign Appeals and the Value of Ambiguity,” *Journal of Politics*, 75(2): 520-534.

The anxiety is all the more understandable because the conduct of Facebook in this and other scandals suggests it has not taken seriously its solemn civic role as a facilitator of news and of political communications.

For this reason, I believe it is critical that independent researchers examine Facebook's data to study the strategies employed in the past by firms like Cambridge Analytica and strategies that more sophisticated firms will employ in the future. Even if Cambridge Analytica was unsuccessful in its attempt to engage in manipulation, new firms will arise that will try harder. Researchers thus need access to specific data on the ads shown to users matched to information that will allow researchers to measure short-term and long-term effects on those users. This will require that researchers utilize sensitive personal data but in a secure context overseen by research ethics boards that protect subjects under study.

Professors Gary King of Harvard and Nate Persily of Stanford Law School have begun an effort to facilitate independent research with Facebook data.⁷ The success of this program will depend on a serious commitment by Facebook to share its data, even and especially in cases that will bring negative press to the company. Evidence showing the ineffectiveness of Facebook targeting may be bad for Facebook's business model. Evidence showing the effectiveness of Facebook targeting may raise real concerns about manipulation. Nevertheless, Facebook's cooperation with this research initiative is essential to the public interest in knowing the power and limitations of online political targeting.

The Supply and Demand of Provocative Political Information

In describing the limits of targeting and likely limits of Cambridge Analytica's efforts in the 2016 election, I do not intend to argue that there is nothing to be concerned about when it comes to the increasing role of social media firms in American democracy. Rather, I intend to help focus attention on issues more important than Cambridge Analytica's targeting strategies. Permit me to raise one such issue here.

Lost in the public conversation about the companies that *supply* political content is the *demand* for such content by politically-engaged citizens. On platforms like Facebook, citizens are sharing and consuming information as a form of *political hobbyism*: they are engaging in politics not out of civic duty but out of a desire for instant personal gratification.⁸ Information, sponsored content included, is shared by users not just to help one another learn about political news, but also to convey a story about themselves and to provoke others. News, both real and fake, is disseminated among users because it feels good to share. The kinds of news and content that often piques our interest appeals to our basest instincts; we are drawn to extremism, provocation, and outrage.

⁷ Gary King and Nate Persily, "A New Model for Industry-Academic Partnerships, Harvard University Working Paper, April 9, 2018.

⁸ Eitan D. Hersh "Political Hobbyists are Ruining the Country," *New York Times*, July 2, 2017.

Facebook's newsfeed is not designed to provide readers with a mix of topics and perspectives that are, in a professional editor's view, important for a citizen fulfilling a duty to be informed. Rather, Facebook's newsfeed facilitates clicks and shares of content – including content supplied by parties, campaigns, and consulting firms – that plays to the appetites of political hobbyists who seek a spectacle.

The same republican principle that demands that political leaders act as intermediaries between popular passions and lawmaking also demand that citizens take cues from editors as intermediaries in news consumption. News readers need news editors. I thus close my testimony by encouraging members of the committee, not just in their roles as legislators but in their roles as civic and political leaders, to direct constituents' attention toward news organizations led by editors who take seriously their duty to inform the public about the range of news and commentary necessary for informed citizenship. At this time, Facebook is not that kind of organization.