SYMPOSIUM ISSUE[†]

YES WE CAN (PROFILE YOU): A BRIEF PRIMER ON CAMPAIGNS AND POLITICAL DATA

Daniel Kreiss*

During the summer of 2011, Michelle Bachman's campaign rolled out an online video advertising campaign exclusively for Republicans likely to caucus living within one hundred miles of the straw poll in Ames, Iowa.[1] In the months leading up to the caucuses Mitt Romney's presidential campaign purchased ads that ran before all YouTube videos watched by voters in Iowa and New Hampshire.[2] Meanwhile, through sophisticated voter modeling, targeted communications based on voters' political interests, and tracking the attitudes of supporters over the course of two campaigns, Romney's campaign orchestrated his near-victory in the 2012 Iowa caucuses.[3]

Underlying all of this is a vast data infrastructure that has made targeted online advertising and marketing possible, and has contributed to a revival of field campaigning over the last decade.[4] Online advertising and field campaign-

---

    † The Privacy Paradox: Privacy and Its Conflicting Values.

    * Assistant Professor, University of North Carolina at Chapel Hill.

    1. Ben Smith & Emily Schultheis, *Targeted Web Ads: The Next Frontier*, POLITICO (Aug. 30, 2011, 2:27 PM), http://www.politico.com/news/stories/0811/62339.html.

    2. Kate Kaye, *Obama Camp Horns in on Iowa GOP Online Turf*, CLICKZ (Jan. 3, 2012), http://www.clickz.com/clickz/news/2134968/obama-camp-horns-iowa-gop-online-turf.

    3. *See* Sasha Issenberg, *Anatomy of a Narrow Victory*, SLATE (Jan. 4, 2012, 11:15), http://www.slate.com/articles/news_and_politics/victory_lab/2012/01/romney_s_iowa_win_i t_took_a_lot_more_than_money_.html.

    4. For a discussion of the corollary of these data practices in commercial contexts, see JOSEPH TUROW, THE DAILY YOU: HOW THE NEW ADVERTISING INDUSTRY IS DEFINING YOUR IDENTITY AND YOUR WORTH (2011). For a history of field campaigning over the last

ing rely on voter modeling based on hundreds of data points culled from surveys, public records, and commercial information sources such as credit histories. This data details the location, demographics, political affiliations, social networks, behavior, and interests of citizens.

In the remainder of this Essay, I discuss the history of political data, focusing on the recent proliferation in voter data and development of new voter-modeling techniques. I conclude with a discussion of the ways these data practices undermine privacy and democratic practice, even as they increase participation and voter turnout.

Gathering and acting on data about the electorate has a long history, but the sheer expanse of data now gathered and stored about the electorate and the modeling and targeted communications it supports are qualitatively new. Both political parties, as well as a host of commercial firms, have amassed enormous national voter databases that they maintain and provide as a service to candidates from mayoral to presidential races. There are a number of different sources of this political data. The core of these databases are public data collected from local, state, and federal records, which include information such as party registration, voting history, political donations, vehicle registration, and real estate records. This data is supplemented with commercial information such as magazine subscription records, credit histories, and even grocery "clubcard" purchases.[5]

These data are continually updated during campaigns through the efforts of thousands of volunteers. In the final two months of the 2008 presidential election, for instance, Obama's millions of volunteer field canvassers gathered over 223 million pieces of information that are now stored in a database owned by the Democratic Party.[6] Parties carry these databases across election cycles and make them available to their candidates running for office at all levels of government, who in turn continually update the voter files.[7]

Data becomes meaningful only through voter modeling. After the midterm elections, consultants for both parties began to more systematically tie these models to actual data on voter attitudes and behavior. Through much of the last decade campaigns were swimming in data that were not tied in any meaningful way to voter attitudes or behavior. What proved effective was distilling hundreds of data points into simple categories of voters: likely supporters, those that can be persuaded, and those supporting another candidate. For example,

---

decade, see RASMUS KLEIS NIELSEN, GROUND WARS: PERSONALIZED COMMUNICATION IN POLITICAL CAMPAIGNS (2012), *available at* http://press.princeton.edu/titles/9616.html.

    5. *See* MARK ANDREJEVIC, ISPY: SURVEILLANCE AND POWER IN THE INTERACTIVE ERA (2009).

    6. David Talbot, *The Demcorats' New Weapon*, TECH. REV. (Dec. 18, 2008), http://www.technologyreview.com/computing/21902/.

    7. Daniel Kreiss, Taking Our Country Back: The Crafting of Networked Politics from Howard Dean to Barack Obama 20 (Aug. 15, 2010) (unpublished manuscript) (on file with author). For more detailed information about the Democratic Party's databases, as well as the modeling detailed below, see generally *id*.

the 2008 Obama campaign hired a consulting firm, Strategic Telemetry, to create its voter models.[8] The firm began by surveying a random, representative sample of the electorate. The firm then looked for correlated data points among Obama's supporters and undecideds. Strategic Telemetry then built models of voters from these combinations of data points and layered them onto the voter file, generating a composite score of likely support for Obama on a zero-to-one-hundred scale for every member of the electorate. The firm then continually polled the electorate and incorporated the results of field canvasses to refine its models.

The Obama campaign targeted priority individuals residing in heavily Republican districts, and focused on neighborhoods with low voter turnout but high numbers of likely supporters. The Obama campaign also developed its online advertising strategy using these voter models and targeting strategies, including using geo-location targeting made possible by IP addresses to display ads to individuals residing in congressional districts with high concentrations of Democratic voters and favorable demographics.

The data that support this modeling are increasingly being merged with information about the online identities and behavior of voters. This means integrating databases that contain information about email list sign ups and "friends" of candidates on Facebook with the large-scale voter files detailed above. While it is difficult to have an accurate picture of an in-progress campaign, it is clear that the integration of databases to deliver more targeted communications is a key theme of the 2012 campaign. The firm Campaign Grid, for instance, is actively matching voter databases, including the Republican Party's, to the online registration data of its multiple partner sites.[9] This allows the firm's clients to deliver video, display, and search advertising to targeted segments of, and even individual, voters.

In two recent pieces, Philip Howard and I identified three ways that this proliferation of political data undermines political privacy and threatens democratic practice.[10] Despite these concerns, institutional political actors that are not official arms of the state, such as parties, candidates, and advocacy organizations, currently enjoy wide latitude to collect and store political data under the auspices of political speech. And yet, as the discussion below suggests, even as these data practices support political participation and mobilization, they come with a democratic cost.

---

8. *Id.* at 280.

9. *See CampaignGrid Announces Technological Breakthrough in Political and Public Affairs Online Advertising*, CAMPAIGNGRID (May 5, 2011), http://campaigngrid.com/blog/CampaignGrid_News/post/CampaignGrid_Announces_Technological_Breakthrough_in_Political_and_Public_Affairs_Online_Advertising/.

10. Daniel Kreiss & Philip N. Howard, *New Challenges to Political Privacy: Lessons from the First U.S. Presidential Race in the Web 2.0 Era*, 4 INT'L J. COMM. 1032 (2010); Philip N. Howard & Daniel Kreiss, *Political Parties and Voter Privacy: Australia, Canada, the United Kingdom, and United States in Comparative Perspective*, FIRST MONDAY (DEC. 6, 2010), http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2975/2627.

First, there is the reality, and future risk, of data breaches and the unauthorized dissemination of sensitive citizen information. Political data is traded on a largely unregulated and international market. Multinational credit firms such as Experian service much of the political sector, and both parties have outsourced the technical development of campaign tools and databases to third parties, including foreign entities. A number of recent incidents reflect the growing potential for abuse of political information. In 2003, journalists from *Wired* purchased data from the political firm Aristotle under assumed names and in violation of state laws, paying $25 per 1,000 voters.[11] During the 2008 general election campaign, the computers of McCain's and Obama's staffers were hacked and data was stolen.[12] While data breaches are a concern in any context, the information held by political actors reveals individuals' policy preferences and political ideology and therefore is particularly sensitive.

Second, privacy is important for protecting anonymous speech and freedom of association. Privacy helps ensure robust political debate by providing citizens the opportunity to form their own viewpoints, craft arguments, and develop political identities free from surveillance and public pressure, all of which also preserves a space for dissent from prevailing social norms. There is little evidence that citizens are becoming wary of the media and its associated privacy concerns, even in the face of the pervasive gathering of information and growing concerns of scholars. But as citizens become more educated about privacy-related issues they may be less likely to engage in public and private political expression.

Third, there are enormous informational asymmetries in the contemporary use of data that implicate political competitiveness, discourse, and representation. Political data and the consulting services necessary to render it actionable are not cheap. Wealthy candidates and those with deep-pocketed allies have a competitive advantage in their ability to purchase comprehensive voter data and sophisticated modeling and targeting services. Minor party and insurgent candidates within parties have comparatively fewer resources to spend on these services. Even the prices parties charge their own candidates to access their voter files can be prohibitively expensive. For example, the Iowa Democratic Party charged its presidential candidates $100,000 to access its data in 2008.[13]

Campaigns also use data to actively and surreptitiously shape political discourse.[14] Campaigns routinely "redline" the electorate, ignoring individuals

---

11. Kim Zetter, *For Sale: The American Voter*, WIRED (Dec. 11, 2003), http://www.wired.com/politics/security/news/2003/12/61543.

12. Lee Glendinning, *Obama, McCain Computers "Hacked" During Election Campaign*, GUARDIAN (Nov. 7, 2008, 9:53 AM), http://www.guardian.co.uk/global/2008/nov/07/obama-white-house-usa.

13. Kreiss, *supra* note 7, at 174-75.

14. The hidden nature of much of these practices fits Daniel J. Solove's Kafka metaphor for privacy, in which campaigns use information about citizens to make decisions about them while denying them any say in how the data are used. *See* Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 477 (2006).

they model as unlikely to vote, such as unregistered, uneducated, and poor voters.[15] Campaign volunteers do not show up on these voters' doorsteps and, if they live in electorally undesirable areas, candidates do not come to their neighborhoods. Furthermore, with the merging of databases, the information environments that these citizens inhabit are increasingly qualitatively different from those of their politically engaged peers, unbeknownst to them. Disengaged citizens are less likely to see political advertising given that campaigns do not spend resources on individuals unlikely to vote.

At the same time, candidates can increasingly choose not only which voters they wish to interact with, but the face of their public selves to present. Quite apart from *who sees a message* is the *content of that message*. Campaigns are increasingly tailoring political communications down to narrowly defined segments of the electorate, and even to individuals, through direct mail, online advertisements, and face-to-face voter contact. This means that campaigns can develop narrow appeals based on ideology and self-interest and direct them to different groups of voters, appearing to be all things to all people.

In conclusion, campaigns use data to expand political participation. Campaigns model the electorate to find their supporters and engage in extensive ground operations and targeted online advertising to fashion them into donors and voters. While this enhanced political participation is normatively desirable on some grounds, it comes with the cost of the erosion of political privacy and democratic debate. Citizens are marketed to and monitored at unprecedented levels, and their personal data are traded on a vast, generally unregulated market. The electorate is carved up on the basis of sophisticated data, meaning many are left out of political communication entirely. Meanwhile, informational asymmetries undermine electoral competition and distort the relationship between candidates and citizens.

---

15. *See* PHILIP N. HOWARD, NEW MEDIA CAMPAIGNS AND THE MANAGED CITIZEN 178 (2006).