BIG DATA AND VOTER SURVEILLANCE

Colin J. Bennett
Department of Political Science
University of Victoria, BC, Canada
Cjb@uvic.ca
www.colinbennett.ca

Paper prepared for presentation to Workshop on Big Data Surveillance,
Queens University, May 12-14, 2016

Recent presidential elections in the United States, and elsewhere, have raised to public attention the general question of how political parties and candidates process and analyze personal data on individual voters. The conventional wisdom, whether accurate or not, is that the modern political campaign needs to be "data driven" to consolidate existing support and to find potential new voters and donors. The capture and consolidation of these data permit the construction of detailed profiles on individual voters and the "micro-targeting" of increasingly precise messages to increasingly refined segments of the electorate, especially in marginal states and districts (Issenberg 2013; Rubinstein 2014). The logic of 'Big Data' has now penetrated electoral politics.

I have argued elsewhere that these practices constitute a form of surveillance. Just as we have conceptualized consumer or employee surveillance, it is logical to isolate and examine voter surveillance, and consider its distinctive dynamics, risks and norms (Bennett, 2015a). Four broad trends are apparent: the move from voter management databases to integrated voter management platforms and integrated campaign "toolkits" for website development, social media strategies, and political messaging; the shift from mass-messaging to micro-targeting employing personal data from commercial data brokerage firms; the analysis of social media and the social graph that allows for peer pressure or "targeted sharing"; and the decentralization of data to local volunteers and workers through mobile applications. In further writing, I have tried to investigate the privacy risks inherent in voter surveillance (2013a, 2013b), and the various policy/legal questions in North America and Europe (Bennett, 2015b).

This brief paper surveys our current state of knowledge about the data-driven campaign, and then poses a series of empirical and theoretical questions that will drive research in this area over the course of the Big Data Surveillance project.

Big Data and Modern Election Campaigns

The political cultures of the United States, and to a lesser extent Canada and Australia, have historically been far more tolerant of a variety of practices to monitor and profile the electorate, and to use direct marketing techniques to poll, canvass and get-out-the-vote. Rubinstein summarizes the current situation in the United States:

> Political databases hold records on almost 200 million eligible American voters. Each record contains hundreds if not thousands of fields derived from voter rolls, donor and response data, campaign web data, and consumer and other data obtained from data brokers, all of which is

> combined into a giant assemblage made possible by fast computers, speedy network connections, cheap data storage, and ample financial and technical resources. Ubiquitous personal identifiers (name and address, telephone numbers, e-mail addresses, IP address, cookies, mobile device IDs, and other unique IDs) allow campaigns to link and integrate these diverse datasets, while data mining and sophisticated statistical techniques allow them to engage in highly strategic and cost-effective analysis and targeting. (Rubinstein, 2014: 861)

These practices are facilitated by the widespread availability voter registration data, largely facilitated by the Help America Vote Act (HOVA) of 2002, passed in the wake of the irregularities and inefficiencies in the 2000 elections. HOVA requires states, among other things, to maintain a "single, uniform, official, centralized, interactive computerized statewide voter registration list."[1] This legislation helped lay the groundwork for political parties to build massive databases of all voters, and also for commercial data brokers to get into the business of compiling, analyzing and selling voter intelligence data (Hersh, 2015).

There are "in-house" databases for both main parties: "Votebuilder" for the Democrats, operated by NGP Van; and the GOP Data Center (formerly Voter Vault) for the Republicans. Both systems are based on state voter registration data, which are then supplemented by a variety of other sources of data from commercial and public sources, as well as from telephone polling and voter contact (Judd, 2013; Howard and Kriess, 2010). Both systems have their origins in the 1990s, but until the HOVA was passed they were incomplete (Hersh, 2015, p. 67).

Perhaps more important than these in-house systems are a number of commercial operations that offer not just databases, but integrated voter management platforms that provide an entire suite of services for any campaign: website design and development; social media outreach; the generation of geo-targeted lists for e-mail and texting; the management of volunteers; as well as the publication of more traditional campaign materials. These platforms also integrate data from commercial data brokerage sources such as Acxiom, Dun and Bradstreet and InfoUSA (Issenberg 2013).

On the Democratic side, the main example is Catalist, best understood as a "data cooperative" according to its chief executive, Laura Quinn (Economist, March

[1] Section 303 of the Help America Vote act (HAVA) at: http://www.eac.gov/assets/1/workflow_staging/Page/41.PDF

26th, 2016, p.6). In the Catalist database, every voter is listed with more than 700 descriptive fields, almost half of which come from commercial sources (Hersh, 2015, p. 169). After the 2012 election, the Koch billionaires invested in a parallel operation called i-360, which has been generating voter data for the candidates in the 2016 elections, and claims a massive database of 190 million registered voters:

> So we've got quantity – but what's even more important to us is the quality of our data. To ensure it is as accurate as possible, we update our data constantly. We source thousands of attributes from multiple consumer data compilers, constantly refresh voter registration information from all states and gather millions of political and issue attributes on an ongoing basis. We then expand the efficacy of this data by using it to build our national predictive models that help clients answer unknowns through the most advanced data science.[2]

It is also worth noting that some campaign organizing companies are agnostic as to the type or ideological purpose of the campaign that they support. So a company like *Nationbuilder*, for instance, now boasts 7000 customers in 98 countries, ranging from Amnesty International to *AirBnB* to the Republican Party of Florida to Arizona State University.[3] One of the oldest companies in the business, *Aristotle.com*, is similarly non-partisan.

However, we should not overstate the value of commercial data to the modern campaign. Popular writing about these technologies, as well as the political consulting business, typically oversells the technologies. In reality, much of the data used by campaigns in the US is quite prosaic and publically available. In recent research on the effects of these databases on campaigning in the US, even the most sophisticated ones like those of Barack Obama, do not have accurate and detailed information about the preferences of voters, and commercial data is often inaccurate, dynamic and not weakly correlated with indicators of political affiliation (Hersh, 2015, p. 176). Thus, "when campaigns perceive voters, they do not see the opinions, traits and behaviors that voters see themselves. They see perceived voters, a simplified and distorted version of the electorate that is based on the data available to them" (Hersh, 2015, p. 12).

Question 1: How are consumer data being used in political campaigns in the United States, and what is the difference between using these data for commercial marketing and political marketing?

---

[2] "What we Do": at: http://www.i-360.com

[3] http://nationbuilder.com

Question 2: How does the logic of 'Big Data' analytics apply to the modern election campaign?

The story is not solely one of the amalgamation and centralization of Big Data to the benefit of central party operations. These trends are offset by the development of campaigning techniques that have harnessed the more decentralizing powers of social networking tools and mobile applications. In recent election cycles, mobile apps have been used for: more traditional one-way political messaging; for door-to-door canvassing; for event management; for encouraging donations; and for broader civic engagement. The website Capterra.com lists over 50 products with a range of features designed to manage election campaigns, grassroots organizing, fund-raising, advocacy, constituency building and so on.[4] Smart canvassing technologies enable the capture of more data elements, and can be combined with publically available demographic data more readily. They facilitate the more widespread lateral sharing of personal information on who votes and for whom, and have the potential to place data on political affiliations, beliefs, and behavior in the hands of ordinary campaign workers and volunteers who may have little privacy and security training (Bennett, 2015b).

Parties in many countries are also becoming increasingly adept at using social media to target messages, recruit volunteers and donors and track issue engagement. Campaign advertisements on social media sites like Facebook are more easily customized to target audiences. Broadcasting is being replaced by "narrow-casting" or "micro-targeting." Social media provide a more effective way to "fish where the fish are," according to one Canadian campaign manager (Delacourt, 2015).

The ideal goal for any campaign is to unleash the real potential for social media as a networking tool. Ideally any campaign would like to have full access to the "social graph" by, for instance, tapping Facebook supporters' social connections and by comparing their "friend" lists with the wider voter databases. NGPVan has pioneered a Social Organizing Application for this purpose: "Social Organizing provides clients with the ability to have supporters match their Facebook friends to the voter file as they take part in everyday campaign activities like voter identification and persuasion, grassroots fundraising, crowd building, volunteer recruitment, and get-out-the-vote activities."

[4] http://www.capterra.com/political-campaign-software/

The analysis of a user's social graph can lead to what has come to be known as "targeted sharing" (Sherer, 2012). In the final weeks of the 2012 campaign, over 600,000 Facebook friends of the Obama campaign signed up for an Obama for America application that allowed the sharing of specific content about the Obama campaign with their friends. In an instant, the campaign had access to more than 5 million contacts that potentially saw each other registering to vote, giving money, sharing videos on the campaign, and voting on or before Election Day. When matched against other voter files, these voters were prioritized for further contact. In reaction to growing privacy concerns, however, it was reported in 2012 that Facebook cracked down on how much information third-party apps could gain about friends' lists. The "Ready for Hillary" campaign, therefore, which initially used targeted sharing, has since become unable to keep up-to-date with supporters' friends lists (Ward, 2014).

A larger shift in campaign logic underlies many of these new trends, namely that voters are more likely to be persuaded if they see their peers supporting a particular party or candidate (Issenberg, 2013). Initial assessments of the Obama for America Facebook app revealed that the click-through rate on a targeted share was more than twice as that for a standard banner ad (Judd, 2012). Some scientific studies have also indicated that this kind of "targeted sharing" through Facebook can have a small but significant impact on voting, especially among the 18-29 age-group (Bond et al., 2012). Again, however, the perceived benefits can be exaggerated. Hersh's research demonstrates that even if friends and family members of supporters can be identified, it is not clear that they would want to engage in a political conversation. Overwhelming his research indicates that the main, and most successful, strategies pursued by the contemporary campaign still rely on volunteer field contact for mobilization and persuasion (Hersh, 2015, p. 190).

Question 3: How are social media used to target and monitor voters? Are these surveillance practices centralizing or decentralizing, hierarchical or lateral, top-down or peer-to-peer?

Big Data and Voter Surveillance outside the United States

Although there are huge differences between presidential and parliamentary systems, there is evidence that parties in other countries are drawing lessons from the American experience, and that similar techniques are gradually entering their politics (Bennett 2013b). There is extensive cross-national communication about these techniques through the network of political and technical consultants, who are eager to tout the benefits of micro-targeting and data-driven campaigning, and to sell the range of software applications, for both database and mobile environments.

Privacy (data) protection law in other countries prohibits this kind of massive capture of personal data on political beliefs and affiliations, the larger parties in some Westminster systems have adopted centralized and internal voter management systems along the lines observed in the U.S. In Canada, there has been close collaboration between Republican consultants and the Canadian Conservative party, whose Constituent Information Management System (CIMS) was developed in 2004 using the Voter Vault software used by the Republicans. The Canadian Liberal Party has a similar "voter identification and relationship management system" called Liberalist, based on the Democrats' Voter Activation Network platform. The left-of-center New Democratic Party uses a system called Populus. There was heightened scrutiny of these systems during the October 2015 general election, and repeated calls from the parties to be covered by Canadian privacy law (Bennett and Bayley, 2012; CBC, 2015; Bennett, 2015b). Similar party databases have been observed in Australia for the last decade or so (van Onselen and Errington, 2004).

The only European country whose parties admit to operating voter management databases of the kind seen in North America is the UK (Amberhawk, 2013; Anstead, 2015). The Conservative Party originally used the "Voter Vault" software developed by the Republicans and then shifted to MERLIN (Managing Elector Relations through Local Information Networks) (Crabtree, 2010). There was a report that it made a further, and quite late, changes for the 2015 election, adopting a new system called VoteSource, which still created confusion among party workers and candidates (Abbott, 2015). These complaints are ironic in the context of the unexpected election of a majority Conservative government, which was partly attributed to a more effective "ground game" and more superior data analytics (Ross, 2015). The Labour Party adopted a system, developed by Experian, called "Contact Creator" in 2008. The system was supposed to

integrate membership lists with voter identification information from the electoral roll, and place this in the hands of local campaigners. The system was retooled using Nationbuilder software in 2013 (Ferguson, 2013). The Liberal Democrats adopted a version of the Voter Activation Network system for the 2015 election (Cookson, 2015).

The internal data processing operations of political parties in every country are typically shrouded in a good deal of secrecy. The inherent competitiveness of the electoral environment, and the proprietary nature of the new campaigning technologies, mean that outsiders have considerable difficulty discovering the extent to which parties might capture data on the wider electorate, beyond that of their members and donors. Nevertheless, it would seem that there is a trend towards the amalgamation of more, and detailed, information on the behavior and preferences of voters, and political consultants are aggressively marketing the benefits of these systems outside the U.S.

Question 4: How, and through what channels, are voter surveillance practices pioneered in the United States spreading to other democratic countries?

Question 5: What is the institutional, legal and cultural constraints that shape how these practices are, or are not, being accepted in other democratic states?

Theorizing Voter Surveillance

It is apparent that this dynamic world of voting surveillance reflects many of the same trends identified in other surveillance studies. The totality of the practices defies easy categorization. The trends are both centralizing and decentralizing at the same time. They produce some staggering potential for the analysis of "Big Data" as well as more localized potential for grassroots mobilization and participation.

I suggest that the existing theoretical literature on surveillance provides only an imperfect tool to grasp these new practices. Mindful that the surveillance literature is driven by empirical work in criminal justice, the workplace, and consumption, the norms, dynamics and consequences of surveillance in this campaigning and electoral context are, and should be, different (Bennett 2015b). The subject is the voter (or potential voter) rather than the suspect, the employee, the consumer and so on). Different subjectivities, we know, dictate different power dynamics, organizational relations, and technological practices.

A crucial task, then, is to try to suggest how voter surveillance is, or might be, different and how this context might produce different judgments about the historical and empirical dynamics, and also about the moral and normative critiques.

Political parties also constitute a different category of organization in most countries. They are neither governmental nor commercial. Neither do they fit easily into that residual category of the "non-profit" sector. They occupy a unique position in democratic practice and perform essential roles in political recruitment, policy development and political socialization and mobilization. They are (still) the mechanisms that define electoral competition and political identification within modern democracies. Although they have embraced many of the techniques of mass communication and mobilization adopted by the private sector, they remain a unique breed of organization.

Question 6: How can we understand voter surveillance according existing concepts and theories in the surveillance literature?

Voter surveillance poses challenging normative questions. If one takes a more neutral view of surveillance, then one would obviously take into account that the public interest that surveillance serves in this context is a democratic one: engagement, voter education, participation, and connection with the political system. Voter management and micro-targeting, according to this interpretation, are just a logical and more efficient way to understand political desires and to tailor campaign messages to more precise categories of electors. Discerning voters' desires is not easy, and data can help political candidates gain a more nuanced understanding of the views of relevant constituencies on more precisely defined questions (Hersh, 2015, p. 206).

A more critical response would contend that the practices surveyed above discourage engagement and deliberation, in favor of the increasing individualization of political space in which we are assumed to have preferences and tastes that only need to be unearthed using the most sophisticated technology to determine what public policies and goods voters "want." Political actors now "shop for votes" and turn them off the political process (Delacourt, 2013). There is evidence in the US that the precise segmentation of the electorate distorts the perceptions of politicians, reducing the portion of the electorate that they need to care about (Hersh, 2015, p. 209).

A further critique might question how extensive surveillance of the political landscape might lead to the differential distribution of public goods. Voter surveillance, therefore, might support a clientilistic politics where precise knowledge of the voters' desires leads directly to reward. Patron-client relations are, of course, common in many developing democracies (Stokes et al. 2013). But they are also reflected in the machine politics of some US cities, and in the system of distributive politics represented by the federal pork-barrel within the US Congress. The manipulation of data, the building of voter profiles, and the treatment of voters as "consumers" permit a more customized message, and perhaps a more customized set of rewards.

Rewards or sanctions? In the era of suspicion about the omniscient national security state (Schneier, 2015), a critical view would also be concerned about the value of such extensive data within the hands of intelligence and law enforcement. The profiling of the electorate along precise ideological lines is anathema to certain countries for precisely those reasons. There is no evidence to data that these proprietary voter management systems have been accessed by the agencies of the national security state. Yet it is exactly this fear that produces such a critical reaction in European societies to these practices, and to a very strong regulation of the processing of personal data on "political affiliation."

Question 7: What is the broader impact of voter surveillance on democratic politics?

In conclusion, therefore, neither the complex and multi-disciplinary literature on surveillance, nor the empirical and normative literature on privacy fully captures the range of fascinating questions that new practices of voter management and engagement entail. These developments are inherently *political*. They speak to a range of theoretical, comparative and empirical questions about the conduct of elections and the behavior of the electorate in different political cultures. They also raise some profound normative issues about the relationship between the watchers and the watched, the government and the government, and the representatives and the represented.

References

Abbott, Paul. 2015. "Don't just blame VoteSource: The Party Needs Constitutional Change," at: http://www.conservativehome.com/thecolumnists/2015/09/paul-abbott-dont-just-blame-votesource-the-party-needs-constitutional-change.html

Amberhawk Training Ltd. 2012. "Could the Conservative Party's Electoral Database breach the Data Protection Act?" Accessed May 20, 2015: http://amberhawk.typepad.com/amberhawk/2013/03/could-the-conservative-partys-electoral-database-breach-the-data-protection-act.html?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+HawkTalk+%28Hawk+Talk%29

Anstead, Nick. 2016. "Use of Data in the 2015 British General election campaign," Colchester, Essex: UK Data Archive.

Bennett, Colin J. 2013a. "Data Point: What Political Parties Know about You," *Policy Options*, 34 (2) (February): 51-53

_____ 2013b. "The politics of privacy and the privacy of politics: parties, elections and voter surveillance in Western democracies," *First Monday* 18 (8) August 5, 2013. Accessed May 20, 2015: http://firstmonday.org/ojs/index.php/fm/article/view/4789

_____. 2015a "Trends in Voter Surveillance in Western Societies: Privacy Intrusions and Democratic Implications." *Surveillance and Society*, Vol 13, No. 3/4 (2015).

_____ 2015b. "Micro-Targeting, Voter Intelligence and Data Protection Law: Can Candidates and Political Parties do in Europe what they do in North America?" Paper presented to Privacy Law Scholars Conference, Amsterdam, October 26, 2015.

Bennett, Colin J. and Robin M. Bayley. 2012. *Canadian Federal Political Parties and Personal Privacy Protection: A Comparative Analysis*. Ottawa: Office of the Privacy

Commissioner of Canada. Accessed May 20, 2015: http://www.priv.gc.ca/information/research-recherche/2012/pp_201203_e.asp

Bond, Robert M. Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle & James H. Fowler. 2012. "A 61-million-person experiment in social influence and political mobilization," *Nature* 489 (September 13th): 295-298.

Canadian Broadcasting Corporation News. 2015. "Federal election 2015: How data mining is changing political campaigns," September 3, 2015 at: http://www.cbc.ca/news/politics/federal-election-2015-how-data-mining-is-changing-political-campaigns-1.3211895

Crabtree, James. 2010. "David Cameron's Battle to Connect," *Wired Magazine*, March 24, 2010. Accessed May 20, 2015: http://www.wired.co.uk/magazine/archive/2010/04/features/david-camerons-battle-to-connect

Cookson, Robert. 2015. "Parties make it personal with tailored messages in election battle," *Financial Times, Politics and Policy, February 17, 2015 at: http://www.ft.com/cms/s/0/ad97068e-b062-11e4-92b6-00144feab7de.html#axzz3kmneOmyk*

Delacourt, Susan. 2013. *Shopping for Votes: How Politicians Choose Us and We Choose Them.* Madeira Park: Douglas and McIntyre.

_____. 2015. "Facebook could change election ad game," *Toronto Star*, February 13, 2015 at: http://www.thestar.com/news/canada/2015/02/13/facebook-could-change-election-ad-game-delacourt.html

The Economist. 2016. "The Signal and the Noise," March 26th-April 1st: pp. 4-6.

Hersch, Eitan. 2015. *Hacking the Electorate: How Campaigns Perceive Voters* Cambridge: Cambridge University Press.

Howard, Philip N. and Daniel Kreiss. 2010. "Political parties and voter privacy: Australia, Canada, the United Kingdom, and United States in comparative perspective," *First Monday* 15 (12) 6 December 2010. Accessed May 20, 2015: http://firstmonday.org/ojs/index.php/fm/article/view/2975/2627H

Issenberg, Sasha. 2013. *The Victory Lab: The Secret Science of Winning Campaigns*. New York: Random House.

Judd, Nick. 2012. "How Obama for America Made its Facebook Friends into Effective Advocates," *Techpresiden*t, November 19, 2012: http://techpresident.com/news/23159/how-obama-america-made-its-facebook-friends-effective-advocates

Ross, Tim. 2015. *Why the Tories Won: The Inside Story of the 2015 Election.* London: Biteback Publishing.

Rubinstein, Ira. 2014. "Voter Privacy in the Age of Big Data," *Wisconsin Law Review* Vol 2014, No 5: 861-936

Schneier, Bruce. 2015. *Data and Goliath: The Hidden Battles to Collect your Data and Control the World*. New York: Norton.

Sherer, Michael. 2012. "Friended: How the Obama Campaign Connected with Young Voters," Time Magazine, November 20, 2012. Accessed May 20, 2015: http://swampland.time.com/2012/11/20/friended-how-the-obama-campaign-connected-with-young-voters/

Stokes Susan, Thad Dunning, Marcelo Nazareno and Valeria Brusco. *Brokers, Voters and Clienteles: The Puzzle of Distributive Politics.* Cambridge: Cambridge University Press.

Van Onselen, Phillip and Wayne Errington. 2004. "Electoral Databases: Big Brother or Democracy Unbound?" *Australian Journal of Political Science* 39 (2): 349-366

Ward, Jon. 2014. "Facebook shutting down a key path Obama used to reach voters," *Yahoo News*, November 17, 2014: https://www.yahoo.com/tech/facebook-slams-the-door-on-political-campaigns-212248365.html