DRAFT UNDER REVIEW DRAFT

Engineering the Public: Big Data, Surveillance and Computational Politics

By Zeynep Tufekci

Abstract: Digital technologies have given rise to a new combination of *big data* and computational practices which allow for massive, latent data collection and sophisticated computational modeling, increasing the capacity of those with resources and access to use these tools to carry out highly effective, opaque and unaccountable campaigns of persuasion and social engineering in political, civic and commercial spheres. I examine six intertwined dynamics that pertain to the rise of computational politics: the rise of big data, the shift away from demographics to individualized targeting, the opacity and power of computational modeling, the use of persuasive behavioral science, digital media enabling dynamic real-time experimentation, and the growth of new power brokers who own the data or social media environments. I then examine the consequences of these new mechanisms on the public sphere and political campaigns.

Keywords: Big Data; Surveillance; Political Campaigns; Persuasive Communication; Privacy

Engineering the Public: Big Data, Surveillance and Computational Politics

The emergence of interactive, networked technologies ~~which shape the current Internet~~ instilled hopes among some observers that ~~this emergent~~ interactivity in the public sphere could help limit or even cure some of the ailments of ~~late~~ modern democracies. In contrast to broadcast technologies, the Internet offers expansive possibilities for horizontal communication among citizens, while drastically lowering the costs of organizing and access to information. It was hoped that these affordances would open new avenues for citizen participation in civic life. The Internet has been a critical tool for social movements around the world; in the United States as well, citizens of diverse backgrounds use it to engage in political activities.

This trend is not unidirectional. The same digital technologies have also given rise to a data-analytic environment that favors power, incumbents, and the technologically adept. The threat arises specifically from increased reliance on *big data, that is*, very large datasets of information gleaned from online footprints and other sources, along with analytic and computational tools. Big data is often hailed for its ability to add to our knowledge in novel ways and to enrich our understanding (Lazer et al., 2009; Lohr 2012). However, big data is rarely examined as a political process involving questions of power, transparency and surveillance. In this paper, I argue that big data and associated analytic tools foster more effective - and less transparent - "engineering of consent" (Bernays, 1947).

Big data's impact on the civic sphere does not happen in isolation; rather, it is the result of multiple intertwined dynamics. **First**, the rise of digital mediation of social, political and financial interactions has resulted in an exponential increase in the amount and type of data available to individuals, i.e. big data. **Second**, emergent computational methods allow political targeting to move beyond aggregated analysis to rich modeling of *specific individuals*. **Third**,

such modeling allows for acquiring answers *about* an individual without directly asking questions *to* the individual, thus opening the door to a new wave of techniques reliant on subterfuge and opacity. **Fourth**, advances in behavioral sciences have resulted in a move away from models of the "rational human" towards more nuanced, realistic, and effective models of human behavior. In combination with the other dynamics outlined here, these models allow for enhanced, network-based *social engineering*. **Fifth**, digital networks enable these methods to be experimentally tested in real time and for immediate deployment, adding a level of previously unfeasible dynamism. **Sixth**, the data, tools and techniques that comprise these methods require access to proprietary, expensive data, and are driven by opaque algorithms, most of which are controlled by a few Internet platforms. These developments give rise to new, potential power brokers.

In combination, the six dynamics of big data concentrate the power to "engineer consent" (Bernays, 1947) in the hands of those with access to data and to the expensive, complicated tools required to use it effectively. (They also bring about significant privacy challenges which have been aptly explored by Kreiss (2012) and Kreiss & Howard (2010) and thus not covered in this paper). These new practices build upon the growing ability of campaigns to use technology to "manage" the electorate (Howard, 2005), a dynamic which has so far been examined in case studies of Barack Obama's campaigns (Carty, 2011; Issenberg, 2012; Kreiss, 2012), probably the most sophisticated examples to date, as well as an ethnographic account of a congressional effort (Nielsen, 2012).

The trends discussed in this paper predate the spread of the Internet, but involve profound qualitative differences. The scope of data, which are collected in an invisible, latent manner,is unprecedented. Previous data collection efforts (for example, magazine subscriptions) required

complicated, roundabout inferences about their meaning and were only crudely useful as broad profiles. On the other hand, big data, when paired with computational modeling, allows for quiet, direct, and more accurate profiling, targeting, and persuading of individuals. It distances political communication from the public sphere and turns it into an increasingly personalized, private transaction. Overall, the impact is not so much like increasing the power of a magnifying glass as it is like re-purposing the glass by putting two or more together to make fundamentally new tools, like the microscope or the telescope, turning unseen objects into objects of scientific inquiry and manipulation.

This paper explores empirically the conceptual foundations of big-data powered *computational politics*. Previous exploration of these technologies and politics includes, most notably, the prescient analysis by Howard (2005), which anticipates and outlines some of the key aspects examined here, especially with regard to new media. This paper addresses additional, intertwined dynamics. The focus of Howard (2005) was on new media, which is but one element in this new data environment, partly because key developments matured only recently. Tools that now seem integral, such as Facebook and Twitter, did not exist a decade ago; big data and user-generated data not fully blossomed.

There has been a flood of applications and research (Lohr, 2012) from fields such as computer science that provide empirical depth but do not take on the political and civic consequences of big data; hence, conceptual theory-building is important. Popular media, on the other hand, rarely goes beyond exploring big data as a hot, new topic and an exciting new tool, and rarely consider issues of control, privacy or power.

While existing case studies concentrate mostly on the Obama campaigns, and while I, too, pick multiple examples from that campaign, this paper is not meant as a study or indictment

of any particular campaign, nor to imply that all the developments outlined here were fully practiced by either Obama 2008, 2012, or any other campaign. (Nor is the goal of this paper to blame any campaign or campaign staffers for trying to win by using all available tools—on the contrary, this is an analysis of the *overall* civic environment as it pertains to rise of big data, not the specifics of any campaign). The strength of empirically-grounded theoretical analysis is to identify important and emergent dynamics as they are playing out, and to lay the groundwork for further empirical exploration.

The aspects of big data and computational campaigns explored here apply in commercial, corporate and political spheres, albeit with different emphases. While the analysis can be extrapolated to other spheres, I focus on the impact on big data on the civic and political sphere with an emphasis on political campaigns, as they help determine structures and forms of governance and are central to questions of civics (Hillygus & Shields, 2009; Kreiss, 2012).

History and Philosophy of Engineering of Consent

The structure of the public sphere is intrinsically connected to the specifics of the technological infrastructure. The Greek amphitheater, for example, allowed the formation of large-scale publics by constructing amphitheaters that would carry sounds to the upper reaches of the seat, but not the other way around. The amphitheater public was constructed as an "audience": one that can hear, but can only meaningfully "speak" when shouting or clapping in unison. Later technologies of broadcast such as print, radio and television, broadened and enlarged audiences while retaining the basic structure of a listening public. The public of the broadcast world could speak, but with a voice that could mostly be heard in the aggregate, through the voting booth, and during special times (elections or large protests), similar to the

way a theater audience is heard during the intermission and curtain, and may only applaud or boo.

The rise of the Internet, it was hoped, would create interactive publics with greater access to data about their government, corporations and politicians as well as a greater means to organize effective collection action (Shirky, 2008). While it is true that the new public sphere is more interactive and less "silent" compared with the past, this is not in contradiction with extension of control by already powerful and centralized forces.

This debate on meaningful participation in governance for a society that is too large for frequent and direct face-to-face interaction—any social organization bigger than a small village or a hunter gatherer tribe—goes back at least to Plato and Aristotle in written records. At its heart, this debate asks whether citizenry–within its gradually expanding, historically-variable definition—can ever be fully equipped to undertake or understand all the complex decisions that are required for governance, and further, what, if anything can keep those with power in check so that they do not assure perpetuation of their own rule.

Plato, famously, called for kings to be philosophers so that they would rule justly for the good of society but not necessarily by being truthful or accountable (Plato, 2012). A modern incarnation of Plato's call for powerful but benevolent "philosopher kings" emerged in the early 20th century Lippmann-Dewey debates (Dewey, 1925; Lipmann, 1927) where Walter Lippmann expressed pessimism at the possibility of a public actually in charge of governance, and argued that the powerful would always be able to manipulate the opinions, beliefs and ultimately, voting behavior of ordinary citizens—they would be "social engineers", in Karl Popper's terms, who manipulated the public to achieve their own goals. John Dewey, however, believed that it was possible to build social and political institutions –ranging from a free press to a genuinely

enriching education—that would expose and counter the manipulations of the powerful and allow for meaningful self-governance by an educated, empowered citizenry.

Though both Dewey and Lipmann worried about the powerful controlling the public, neither had experienced the full force of broadcast media, yet to come. Broadcast media altered dynamics of politics in fundamental ways. Public-relations pioneer Edward Bernays explained the root of the problem in his famous "Engineering of Consent" article (Bernays, 1947) where, discussing the impact of broadcast on politics, he argued that the cliché "the world has grown smaller" was actually false. The world is actually much bigger and today's leaders, he pointed out, are farther removed from the public compared to the past. The world feels smaller partly because modern communication allows these leaders, potent as ever, to communicate and persuade vast numbers of people, and to "engineer their consent" more effectively.

Bernays saw this as an unavoidable part of any democracy –he believed, like Dewey, Plato and Lippmann had, that the powerful had a structural advantage over the masses. However, Bernays argued that the techniques of "engineering of consent" were value-neutral with regard to message. He urged well-meaning, technologically and empirically-enabled politicians to become "philosopher-kings" through techniques of manipulation & consent engineering.

> The techniques can be subverted; demagogues can utilize the techniques for antidemocratic purposes with as much success as can those who employ them for socially desirable ends. The responsible leader, to accomplish social objectives, … must apply his energies to mastering the operational know-how of consent engineering, and to out-maneuvering his opponents in the public interest. (Bernays, 1947, p.115)

Bernays recommended study of the public through opinion research and controlling it mainly through managing of communication and media. The techniques of opinion control

espoused by Bernays became bread-and-butter of political campaigns in the post-war Western democracies. At its heart, this has been driven by "public opinion research" which seeks to understand, categorize and label the public as differentiating is key to effective "persuasion"—whether marketing a politician or a soft drink. Soon after public opinion research started seeping into politics, cultural critic Adorno called the forms of "classifying, organizing and labeling" as a form of propaganda in which "something is provided for all so that none may escape." (Adorno & Horkheimer, 2002, p.123).

However, messaging and mobilization based on categorization has limits intrinsic to the method as all categorization hides variation. The match between demographics or political profiles and a specific person is, at best, broadly probabilistic and often very muddled. During the broadcast era, most targeting was necessarily course-grained, because TV audiences were measured in broad demographics. The best that aspiring micro-targeters could do was to define potential segmented audiences, like "soccer moms", by gender and age, and target programs to gender and age group. Because audiences could not be tightly defined, messaging had to be broader as well. Many exposed to the ads would not fit the target group, and many members of the target group would be excluded. Research showed that such political advertisements on broadcast TV remained largely ineffective in tipping the scale between existing candidates, at least when compared with more structural factors such as the unemployment rate or economic growth. (The political advertisement climate, and the need to advertise on broadcast, arguably, has a stronger effect in determining who can be a candidate in the first place, and not so much in selecting a winner among those who make it to that level).

Similarly, almost all voter canvassing and turnout campaigns have traditionally been based at the precinct level simply because demographic data has been available at that level.

However, precinct data are probabilistic in that no precinct uniformly votes for a single party, so campaigns tend to pour resources into a specific precinct in the hopes that they will mobilize more supporters than opponents, and that their canvassing efforts will not aggravate too many supporters of the other party (Issenberg, 2012). They, in turn, ignore precincts which contain many of their own voters, but less than those of their opponent.

Unsurprisingly, targeting individuals *as* individuals rather as members of broadly defined aggregates has long been the holy grail of political campaigns. Such efforts have been underway for decades. Culling information from credit cards, magazine subscriptions, voter registration files, direct canvassing efforts and other sources, political parties, as well as private databases, have compiled as much information as they can on all individual voters. However, until recently, the collection of individual level data was messy and fragmented, and targeting was still on done by aggregate groups, which were simply based on richer individualized data than before.

New Dynamics of Persuasion, Surveillance, Campaigning and Social Engineering

The recent rise of big data and computational techniques are changing political communication with citizens once again. If the 20th century engineers of consent had magnifying glasses and baseball bats, those of the 21st century have acquired telescopes, microscopes and scalpels. In this section, I examine six, intertwined dynamics which create a new environment of surveillance and social engineering.

1. Big Data: The advent of digital and networked technologies has caused an explosion in the amount and variety of data available on each individual, as well as the velocity with which such data become available. A significant portion of social, political, financial and civic

interactions and transactions that were ephemeral, momentary and lost to time without a permanent record are now mediated through digital environments making it possible to harvest their imprints and thus resulting in a large amount of individualized information that would simply be unthinkable to collect in the pre-digital era. This has resulted in a qualitatively and quantitatively different data environment that can be harnessed for political campaigns (as well as for other purposes). These large collections of data, referred to as big data, is not just more of old kind of data; rather, in some ways, its effects are like the invention of the microscope (Brynjolfsson, 2011) which makes visible the previously existed unseen, and in other ways, like a telescope that allows the observer to "zoom out" and observe at a different scale, perhaps at loss of subtlety and individuality of the data points. While no single metaphor fully captures its novel impacts, big data, like the microscope and the telescope, threatens to upend our understanding of multiple fields and to transform the practice of politics.

It's not just that the depth and the scope of the available data have changed with big data; the fundamental nature of data available for aggregation has undergone a significant shift. In the past, data collection was primarily "pull" (questions answered voluntarily as in surveys), supplemented by a layer of "latent data," which are data which exist as imprints of actions we carry out as we go about our lives In the pre-digital era, such *latent*, imprint data was limited—financial transactions, magazine subscriptions, credit-card purchases. Political campaigns were faced with the task of *inferring* what such a transaction met. Does a subscription to "Better Homes & Gardens" imply a party affiliation? Does it correspond to a position on progressive taxation? The answer often was, maybe, but only weakly. Such data provided some correlational guidance at the group level but did not allow precise individual targeting

The rise of the Internet itself as a social, civic and political space engendered a

tremendous growth in a different category of data often called "user-generated" data. Some of this growth is of *latent* data; transactions which are carried out for a variety of purposes now leave behind harvestable imprints. The part that is of most interest to political campaigns is data I define as "semantic interaction with imprint", which are actual conversations, comments and other relevant interactions that the person engages as a part of her civic and social life. Hence, rather than being asked by a pollster about her choices, and the answer recorded on a Likert scale or another kind of multiple choice format, campaigns can now capture actual utterances of people as they talk about their preferences in politics and related issues. Much such user-generated content is directly semantic, and rather than convoluted inferences, such data lends itself to deeper and direct insight into a person's opinions, dispositions and behaviors (through computational methods discussed below). These new data sources provide more declared, latent *and* semantic data than a political pollster in the twentieth century could ever hope for.

The user-generated content environment has undergone such a dramatic change that even a mere eight years ago, when the Internet was already widespread, political campaigns had to resort to some degree of trickery to compel users to provide content. Howard (2005) documents how a political company operated a discussion forum that provided the participants with the aggregated voting record data of politicians *primarily so they could have access to the participants' discussions*. People had to be coaxed into user-generated data. These days, such data is voluntarily and widely generated by individuals themselves as a by-product of digitally-mediated participation in civic life.

Further, the quantitative depth of big data composed of online imprints is exponentially richer than pre-digital data. A big commercial database may contain from 100 to 1500 data-points on each individual. A person can have that many tweets in just a month, let alone the data

generated from the websites they visit, the links they do and don't click on, comments and status updates on various platforms. The volume and variety of big data is qualitatively different. If anything, the problem of data analysis today is data that is too much, too deep, and too varied. However, rise of computational methods and modeling is quickly catching up to the challenge of turning this deluge of data into operationalizable information at the hands of political campaigns.

2. Emergent Computational Methods: The emergent methods in computation are a result of multiple factors. First, there have been new developments in the storage and access of large amounts of data without which much of the computations on big data could not be carried out. Second, new methodologies allow processing of *semantic* information contained in user-generated natural language outputs such as conversations as opposed to already structured data such as a financial transaction. Third, the spread of network analysis methods and tools means that human interactional data can now also be seen through a structural lens through social network analysis. Fourth, the scale of the data allows for new kinds of correlational analyses that would have been difficult to impossible previously.

First, given the amount of data that is being generated, even mere storage has been a challenge. Youtube has 72 hours of video uploaded every minute. As of last year, Facebook was processing about 2.5 billion pieces of content, 2.7 *billion* "like" actions, 300 million photos and, overall, 500 terabytes of data *everyday* (Constine, 2012) (In contrasts, an online depository of books by leading large research libraries in the world contain a mere 78 terabytes of information in total (Anderson, 2008)). Handling such large datasets has recently become easier through the development of techniques like "Hadoop clusters" which provide a system of shared storage along with "Map Reduce" which provides the analytic layer allowing for reliable and quick access to such large datasets. MapReduce is especially useful for political analysis since it is a

"good fit for problems that need to analyze the whole dataset in a batch fashion" which means that a large datafile can be compiled and written once and then accessed many times with speed and ease rather than older database systems like those of financial accounts in a bank which may need to be updated dynamically, and thus were too slow and difficult to use for analyzing recent big datasets. Facebook reportedly holds its data in a 100-petabyte Hadoop cluster.

Second, new computational processing techniques allow for extracting information from data without using an army of human coders and analysts, as would have been required under old techniques. Techniques of automatically "scoring" words to generate estimates of ideological content of texts and sentiment analysis allows for a probabilistic but fairly powerful method of categorizing a person's approach to an issue as represented in her textual statements, but without the costly step of a human reading the actual content. Topic modeling techniques such as "Latent Dirichlet Allocation" (LDA) organize large amount of texts automatically into "topics", themes from texts that can be used to identify the structures and issues of concern of political constituencies. Without these computational techniques, the texts would have to be read and summarized by a large number of people; even then, just aggregating the results would pose a challenge. Algorithms come with pitfalls and limitations; however, they can be useful in providing information that would otherwise be prohibitively costly or impossible.

Third, social network analysis, the roots of which go to sociology in the 1950s, has seen greatly broadened utility and technical expansion in allowing analysts not just map people's views, but also to situate them within social networks. The broadened utility has occurred partly because data that is in the form a network has increased significantly due to online social network platforms that are used for a variety of ends, including politics (Howard & Parks, 2012). Previously, gathering social network information from people was a difficult and costly endeavor

and various biases and difficulties in recalling social network information led to great many difficulties as even small social networks required hundreds of interviews where people were expected to name dozens if not hundreds of social ties. Understandably, such research has always been very difficult and carried out only on small samples.

With the advent of networks that were encoded by the software, network analysis became possible without the difficult step of collecting information directly from people. Researchers also started applying network analysis to broader topics where the "connections" could be interpreted as "links" in a network--such as blogosphere (with each link constituting a network connection) or citation networks of scholars. Network analysis allows identifying various structural features of the network such as "centrality", clustering (whether there are dense, distinct groupings), bridges that connect clusters and much more, which provide very valuable political information in deciding how to target or spread political material. (For example, people with high centrality are useful propagators of information and opinions). Using social network analysis on large datasets, researchers can analyze information diffusion through complex networks and use their findings to more effectively target and tailor their persuasive communication.

Fourth and finally, researchers can now look for "correlations" in these huge-datasets in ways that would have been difficult to impossible before. Such powers, obviously, can also be used to great good. For example, researchers have started identifying drug interactions by looking at Google searches of multiple drugs matched by symptoms—a feat that simply cannot practically be done any other way which would mean surveying all users of all drugs about all side effects.

3. Modeling: Modeling in this context is the act of inferring new information through

analysis of data based on a rich understanding about the relationship between the underlying data and target information. Modeling can be vastly more powerful than "profiling." Profiling within the political campaign context is often limited to the act of attempting to categorize a user by combining available data and often crucially depends on data gather by contacting the person directly. Someone who answered a survey question on environmental issues as "very important to her", for example, is likely an environmentally-conscious voter. Combined with purchase data ("shops at Whole Foods") a campaign may profile her as an environmental voter.

However, the advent of rich datasets that contain imprints of actual behavior and social network information –social interactions, conversations, friendship networks, history of reading and commenting on a variety of platforms–along with advances in computational techniques means that political campaigns (and indeed, advertisers, corporations and others with the access to these databases as well as technical resources) can model *individual* voter preferences and attributes at a high level of precision, and crucially, often without asking the voter a single direct question. Strikingly, the results of such models may match the quality of the answers that were only extractable via direct questions, and far exceed the scope of information that could be gathered about a voter via traditional methods.

For example, a recent paper shows that merely using Facebook "likes" is sufficient to model and accurately predict a striking number of personal attributes including "sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender." (Kosinski, Stillwell, & Graepel, 2013) While parts of this knowledge may seem trivial since some of these, such as age and gender, are traditional demographics and are usually included in traditional databases, it is important to note that these are being "guessed" through modeling, without any prior information

and are not asked or observed from the user, with the implication that such attributes can also be modeled in platforms where anonymous or pseudonymous postings are the norm.

These models of just Facebook "likes" correctly discriminated whether the Facebook user is heterosexual or not in about 88% of the cases; can predict race with about 95% of the time and political party affiliation about 85% of the time (Kosinski, Stillwell, & Graepel, 2013). In other words, just access to Facebook to be processed through a computational model allows for largely correctly delineating Republicans and Democrats without looking into any other database, voter registration file, financial transactions or membership in organizations. Hence, access to Facebook likes plus any other reasonable level of data –available for purchase in the open market—likely corresponds to almost 100% coverage of key traits ranging from demographics to psychographics of a majority of voters in the United States.

Modeling allows access to psychological characteristics that were beyond the reach of traditional databases, as invasive as those might have been considered. Personality traits such as "openness" or "introversion" or "neuroticism" are traditionally measured by surveys, which have been developed and validated by psychologists and used on a large number of people for decades. While such traits themselves may lack nuance and subtlety, they are significantly more nuanced than the crude demographics employed by political campaigners ("soccer moms."). The PNAS study shows that models based on Facebook likes were as good as the scientific scales.

In other words, without asking a single question, the researchers were able to model psychological traits as accurately as a psychologist administering a standardized, validated instrument. Given that Facebook data have been used to accurately model other attributes ranging from mortality rates for hospitals to obesity rates, and given that Facebook is just one facet of information available for big data modeling, it is clear that political campaigns can have

a much richer, more accurate categorization of voters, and without and before necessarily having knocked on their door a single time to ask a single question.

To understand why this is a major shift, consider how different it is compared with cruder, more basic profiling that has been used in traditional survey research to identify "likely" voters—a key political concern to campaigns. For decades, traditional polling organizations and campaigns have been trying to model "likely" voters in their surveys with varying degrees of success. Campaigns do not want to spend resources on people who are unlikely to vote and pollsters need this data to weigh their data correctly. Asking the question itself ("are you likely to vote?") has not proven that useful, due to the well-known socially-desirable answer bias: many "unlikely" voters declare their intention to vote. Previous voting records are also tenuous predictors—besides, there are many young voters entering the rolls. Gallup, whose likely voter model had long been considered the gold standard, asks a series of seven questions that include intent, knowledge (where is your voting booth?), past behavior, and implicit measures (how often do you think about the election). However, even with decades of expertise, Gallup has been missing election predictions badly enough that it became a punch line for jokes. At the White House Correspondents dinner, President Obama made a predictable joke followed by a joke about who didn't see the joke coming: "Show of Hands? Only Gallup?" (Yagoda, 2013) Reports indicate that the reason Gallup failed as badly did was because its likely voter profile failed, and led it to an erroneous prediction of the electorate and incorrect weights.

In contrast, during the 2012 election, the Barack Obama campaign the developed a fairly sophisticated model of "likelihood of turnout" based on its datasets, which generated an index from 0 (not going to vote) to 100 (will certainly vote.) This resulted in a targeted, highly efficient persuasion and turnout effort (which focused mostly on turning out voters that were already

Obama supporters rather than spending a lot of effort persuading voters who would not end up voting). This left the Romney campaigns, reliant on more traditional efforts, so far behind that after their loss, Romney staffers were left exclaiming that the Obama campaign turned out voters that the Romney campaign "never even knew existed" (Rutenberg & Zeleny, 2012).

Finally, big data modeling can predict behaviors in subtle ways and more effectively oriented toward altering behavior. For example, for years, the holy grail of targeting for commercial marketers has been pregnancy and childbirth, as that is a time of great change for families, resulting in new consumption habits which can last for decades. Previously, retailers could look at obvious steps like creation of a baby registry; however, by then, the information is often already public to other marketers as well, and the target is well into the pregnancy and already establishing new consumption patterns. By combining rich data with predictive modeling, however, the retail giant Target is not only able to identify potential pregnancy very early on, it can also estimate the due date "within a small window" so that it can send coupons and ads timed to the stages of pregnancy and parenting. In a striking example, Duhigg (2012) recounts the tale of an angry father walking into Target, demanding to see the manager to ask why his teenage daughter was being sent advertisements for maternity clothing, nursery furniture and baby paraphenelia. The manager, it was reported, apologized profusely, only to receive an apology himself when the father went back home to talk with his daughter, who was, indeed, pregnant. Data modeling ferreted out facts that a parent did not know about his own minor child living under his own roof.

4. Behavioral Science: The predictive analytics that are possible due to emergent big data and computational modeling would not be as valuable without a corresponding rise in sophistication of behavioral science models of how to persuade, influence and move people to

particular actions. Developing deeper models of human behavior is crucial to turning the ability to look, model and test big data into means of altering political behavior.

The founder of public relations, Edward Bernays, himself had posited that people were fundamentally irrational. However, the rational and "utility-maximizing" aspects of human behavior have long been emphasized in dominant academic literature, especially in fields such as economics and political science. The Habermasian (Habermas, 1991) ideal of the public sphere, where status-free actors carried out rational conversations based on merit, also dominated as normative structure even if it was recognized by political campaigners as wishful thinking during craft of politics. While political practitioners may have recognized that the "rational voter" model did not correspond to their experience in the world, there was fairly little systematic analysis of "hooks" for steering such irrationality into desired outcomes.

All this started changing in academia with theoretical developments that emphasized the non-rational aspects of human behavior, and with attempts to measure and test behavior modification within political context. Just as behavior analysis got more sophisticated, for the first time in modern political history, an influx of academics from the behavioral sciences started moving into practical politics, starting with the 2008 Obama campaign. Hence, as recounted by Issenberg (2012), there was a significant shift from "grant narratives" and tidy conclusions produced by pundits towards an operational fight focusing on altering the behavior of individual voters, aided, crucially, by insights from in-house behavioral scientists.

It wasn't that behavioral science could overcome the structural conditions of a campaign like a bad economy or a profoundly unattractive candidate. Increasingly, however, elections are fought at the margins in part because of pre-existing polarization and low turnout; in such a circumstance the operational capacity to find and convince just the right number of individual

voters becomes increasingly important. In the United States, this is exacerbated by the "winner-takes-all" electoral college model which greatly raises the stakes for identifying and turning out the right voters in the right states. (The number of votes that needed to flip to change outcome in the 2012 presidential election was a mere 400,000 in the right states (author calculation)).

While the contribution of behavioral science remains within the nature of politics as usual, its impacts can be significant. For example, researchers found "plain white envelopes" work better than glossy mailings in signaling credibility—and hence were increasingly used by the Obama campaign. A whole different level of targeting, however, can be obtained by merging "psychographic" profiles, obtained from computationally modeling big data, with empirically-based behavioral science. For example, research shows that when afraid, certain categories of people tend to become more conservative and vote for more conservative candidates. By modeling *individual* psychologies through computational methods applied to big data, a political campaign hoping to garner votes of a conservative candidate can plausibly (and relatively easily) identify voters that were more likely react to fear by voting conservative and target them individually through "fear-mongering" tactics designed for their personal weaknesses and vulnerabilities (for example, scares about children for parents; about safety to people who've suffered from accidents; and terrorism, health, or petty crime) while bypassing individuals on whom fear-mongering would not have the desire effect, perhaps even the opposite of the desired one.

To clarify, it's true that campaigns have often resorted to fear (or other tactics that appeal to the irrational) such as the infamous "daisy/nuclear war" ad campaign during the 1964 presidential campaign. However, this can now be done with a scalpel, at an individual level, rather than hitting a whole population with a message that, while it may work for some, will also

certainly generate a backlash that mitigates its effectiveness.

5. Experimental Science in Real Time Environments: The online world has opened up the doors to real-time, inexpensive and large-scale testing the effectiveness of persuasion and political communication. Much campaigning in the past was directed by "tacit knowledge," "gut feel," and deference to traditional expertise and punditry (Issenberg, 2012). Empirical discussions about politics would, at most, focus on surveys and there has been surprisingly little testing or experimentation in political campaigns (Gerber & Green, 2000).

There are a multitude of reasons for limited political experiments including the fact that campaigns are also businesses--and the ecology of consultants who make a living by running campaigns by "gut feel" were never warm to experimentation which might devalue their own strategies and hurt their bottom-line. Campaign operatives, organized by method—the "direct mail" specialist, the field operator—resisted testing which might have shown a method to be less effective (Issenberg, 2012). Field experiments are costly and time-consuming—money and time being the resources on which political campaigns already place the highest premium.

In spite of these obstacles, some experiments were conducted; however their results were often published too late for the election in question. The field experiments conducted in 2001, demonstrating that face-to-face canvassing was most effective for turnout, were published three years later (Green, Gerber, & Nickerson, 2003). These experiments increased awareness that many methods that campaigns traditionally spent money on (for example, slick mailers or phone calls) were not very effective. The cultural shift in emphasizing metrics came fully of age with the 2008 and 2012 Obama campaigns which were notable for their "data-driven culture"—campaign manager Jim Messina declared as early as 2011 that the 2012 Obama campaign was going to be "metric driven" (Scherer, 2012). A culture of experimentation was encouraged and

embraced.

This shift, however, wasn't just a change in outlook but a change in technical infrastructure as well as rise of digital platforms in the practice of politics that, in combination, allowed incorporating real-time experimentation into the very act delivery of the political message. Sometimes called "A/B" testing in homage to its origins in Silicon Valley, which uses it extensively in platform and ad design, this method involves creating multiple versions of a screen or a message to be delivered separately to randomly selected control groups. The results are measured in real time and quickly integrated into the delivery as the winning message becomes *the* message. Methodologically, of course, this is traditional experimental science but it has become possible because parts of the campaign now take place over a medium that allows for experimental affordances: cheap delivery of messages, immediate measurement, ability to randomize recipients, and quick turn-around of results which can then be applied to the next round. Campaign operatives involved in such "metric-driven" campaigns report that their "gut feel" was often shown to be wrong by these experiments (author interviews).

The Obama campaign had incorporated experiments into its methods as early as 2007. For example, in December 2007, when the Obama campaign was still in its early stages, the campaign created 24 different button and media combinations for its splash page (the first page that the visitors land on). Each variation were seen by 13,000 people—an incredibly large number for running a field experiment by old standards, but a relatively easy and cheap effort in the digital age (Siroker, 2010). In the end, the winning combination (showing a picture of his family and his children) had a 40% higher "sign-up rate"—translating it to the total number of people who signed up, the impact may have been as high as an extra 2,888,000 people who signed up (though, of course, that is a maximum effect). Considering that the average

contribution was $21 and that 10% of people who signed up volunteered, the difference would be an additional $60,000,000 and 288,000 more volunteers that came through a cheap, massive and immediate experiment. Through such experimentation, the Obama campaign was led to predominantly feature his family in much campaign material.

Such an on-the-fly, complicated, "large N" true randomized experimentation has been traditionally rare in social sciences, let alone political campaigns, due to costs, efforts and ethical considerations. Digitally-mediated civics provides a means through which political campaigns can now carry out such experiments with ease and effectiveness.

6. Power of Platforms and Algorithmic Governance: Much political and civic speech in happens in the "fifth estate" composed of blogs, micro-blogs and online social media and social networking platforms, a great many of which are privately-owned corporations with opaque policies and algorithms with regards to content visibility, data sharing and many other features of political consequence. These platforms allow for individualized delivery of messages (each Facebook user could see a different message tailored to her as opposed to a TV ad that necessarily goes to large audiences), and are thus quite powerful addition to networked campaign arsenal. However, they are also governed by opaque and proprietary algorithms.

For example, Twitter, an emergent platform that plays a significant role in information sharing among politicians, journalists and citizens, selects and highlights 10 "trending topics" per region, which then gain visibility as they are advertised on the platform itself. The algorithm that selects topics, however, is proprietary which has led to political actors wondering if they were censored (Lotan, 2011) while others try to "game it" by reverse engineering it (Tufekci, 2013). Such algorithms chosen and known only to the platforms have certain biases in how they determine visibility of content that, in turn, is of great political significance.

Further, the kind of "big data" that campaigns desire most often reside within these platforms, and are thus private property. Campaigns can access this data either through favorable platform policies which grant them access to user information--for example, political "apps" such as those of the Romney and Obama campaign can acquire user information as well as information on users who are friends of the original individual who accepts a campaign app, a potential privacy issue campaigns can bypass. Platforms can make it easier or harder for political campaigns to reach such user information, or may decide to package and sell data to campaigns in ways that differentially empower the campaigns, thus benefit some over the others.

Further, a biased platform could decide to use its own store of big data to model voters and to target voters of a candidate favorable to the economic or other interests of the platform owners. For example, a study published in Nature found that civic "go vote" messages that were targeted in Facebook through user's social networks through a voting encouragement app that was deployed by Facebook resulted in statistically significant increase in voter turnout among those targeted, compared with a similar "go vote" message that came without such embedding in social ties (Bond et al., 2012). A platform that wanted to manipulate election results could, for example, model voters who were more likely to support a candidate it preferred and then target a preponderance of such voters with a "civic" message narrowcast so that most of the targets were in the desired target group, with just enough thrown in from other groups to make the targeting less obvious. Such a platform could help tilt an election without ever asking the voters whom they preferred (gleaning that information instead through modeling, which research shows is quite feasible) and without openly supporting any candidate. Such a program would be easy to implement, practically undetectable to observers (since each individual only sees a portion of the social media stream directed and nobody sees the totality of messages in the whole platform

except the platform owners), easily deniable (since the algorithms that go into things like Facebook's news feed are proprietary and closely-guarded secrets), and practically unprovable.

A similar technique could be possible for search results—ordinary users often never see pages that are not highlighted in the first page of Google results and researchers already found that a slight alteration of rankings could help impact an election—without voter awareness (Epstein & Robertson, 2013). Indeed, based on randomized experiments, Epstein & Robertson (2013) concluded that "with sufficient study, optimal ranking strategies could be developed that would alter voter preferences while making the ranking manipulations undetectable." By holding on to the valuable troves of big data, and by controlling of algorithms which determine visibility, sharing and flow of political information, the Internet's key sites and social platforms have emerged as unscrutinized but important power brokers of networked politics.

Consequences and Power of Big Data Analytics

Big-data driven computational politics engenders many potential consequences for politics in the networked era. In this section, I examine three aspects: deep and individualized profiling and targeting; opacity of surveillance; assault on (idea) of a Habermasian civic sphere.

First, the shift to tailored, individualized messaging based on psychographic profiling obtained through modeling brings potential for significant harms to civic discourse. Howard (2005) and Hillygus and Shields (2009) had already presciently warned of the dangers of data-rich campaign. As Howard (2005) crucially argued, the ability to model out unlikely or unpersuadable voters means that the current strategy of focusing presidential politics on "swing states" can now be implemented at the individual level. A house judged as "non-voter" can be skipped while the next one is drowned in campaign material, quietly— thus introducing a new form of categorical inequality into the public sphere. Previously inefficient data-collection and

modeling made such "redlining" difficult and confined it to precinct and state levels. The advent of big data removes a "beneficial inefficiency" (Karpf, 2012) that aided the public sphere.

Messaging will also likely fracture further, encouraging further deployment of potent "wedge" issues at the expense of broadly engaged topics. Campaigns have long tried to use "wedge" issues - issues that are highly salient and important to particular segments of the voting population - such as abortion or gun rights. However, these are a double-edged sword for campaigns in that they elicit significant passion on both sides. Hence, campaigns aim to put wedge issues in front of sympathetic audiences while hiding it from people who might be motivated in the other direction (Hillygus and Shields, 2009; Howard, 2005). Until now, the ability to do just that has so far been limited by availability of data (finding the exact wedge voter) and means to target individuals (Barocas, 2012).

The use of "wedge issues" in direct mail is a telling example and a taste of the power of targeting capacities. Hillygus and Shields (2009) demonstrated that even by 2004, the use of "wedge issues" by campaigns was significantly more prevalent in direct mail, which is only seen by the recipient, compared with broadcast. Further, the opacity of individualized targeting through digital networks creates a new type of "dog whistle" politics, whereby the campaign emphasizes a provocative position only to sympathetic audiences, while it remains invisible to others. Prevalence of wedge issues is further damaging in that it allows campaigns to remain ambiguous on important but broadly relevant topics (economy, education) while campaigning furiously (but now also secretly) on issues that can mobilize a small but crucial segments.

Further, the construction of "wedges" need no longer pertain merely to issues; it can also incorporate psychographic profiles *modeled* from online social data—again, without directly obtaining data by asking the individual. Then, for example, fear-mongering messages can be

targeted only to those motivated by fear. Unlike broadcast, such messages are not visible to broad publics and thus cannot be countered, fact-checked or otherwise engaged with in the shared civic sphere the way a provocative or false political advertisement on broadcast might have been. This form of big data enabled computational politics is a private one, at its core opposed to the idea of the civic space as a public, shared commons. It continues a trend started by direct mail and profiling, but with exponentially more data, new tools and more precision.

The second negative development pertains opacity of these new methods. The current surveillance environment has been compared to the "Panopticon"—Jeremy Bentham's model of a prison in which a guard is able to see all cells without revealing which one he is looking at any moment, and later used as a metaphor for modern surveillance by Foucault (1977). While the observational aspect is similar, big data analytics is currently exercised in a manner opposite of the Panopticon. Whereas the Panopticon operates by making very visible the act and possibility of observation while hiding actual instances of observation, so that the "prisoner" never knows if she is being watched but is always aware that she could be, modern social engineering operates by making surveillance as implicit, hidden and invisible as possible, and, hopefully, without the observed person being aware of it (while the latest NSA revelations due to leaks by ex-NSA employee Edward Snowden may change that, the level of surprise and outrage they generated speaks to both lack of awareness of surveillance as well as efforts to keep it quiet).

While browsers, cell phone companies, corporate and software companies, and, as recently revealed, the US government, accumulate an extensive information about individuals, the depth and the scale of the accumulated data remains opaque and inaccessible to the ordinary person. That the "guard" in such cases, *contra* Bentham, hides the fact of his observation from the prisoner flows from the fact that we are not actual prisoners but rather citizens who may be

upset about surveillance and loss of privacy—and take action against it. As we are not prisoners, the model of control sought by these systems is not that of pure fear as in George Orwell's 1984, another famous work on surveillance and control, but rather an infrastructure of surveillance (and targeted fear aimed at "underclass" subgroups) along with direct overtures toward obtaining assent and legitimacy through tailored, fine-tuned messaging—a more Gramscian model of hegemony. Combined with model information obtained without asking, political campaigns are moving to the paradigm of the modern marketing, which aims to appear as a "pull" system, in which the user seeks and views the ad because it is enjoyable and cool (Super Bowl or popular ads such as those of Chipotle which people willingly watch on Youtube), whereas the actual effort is actually a highly expensive "push" to design the most "pullable" product.

Research shows that people respond more positively to messages that they do not perceive as intentionally tailored to them, and that overt attempts are less persuasive than indirect or implicit messages. Indeed, political campaigns are acutely aware of this fact. As advisor and consultant to the Democratic party, Hal Malchow, who designed campaign messages that looked like they came from more neutral sources and toned down and intentionally made direct mailings look cheaper, puts explicitly: "People want information, they don't want advertising. When they see our fingerprints on this stuff, they believe it less." (Issenberg, 2010).

A third way in which big data driven computational politics can undermine the civic experience is the destruction of "status-free" deliberation of ideas on their own merit, as idealized by Habermas (1991). Habermas' ideal public sphere envisioned a public composed of interactions between status-free individuals where ideas were debated on their merits, regardless of who uttered them. Regardless of whether one thought this ideal existed at all, and taking into account of its critics (Fraser, 1997), the new developments constitute an anti-Habermasian public

sphere in which every interaction happens between people who are "known quantities", and further, the public is constituted unequally as the campaign knows a lot about every person it is interacting with whereas ordinary members of the public have no such information. In the new information environment, "status-free" deliberation is removed from the equation.

Even when identity information is not embedded onto the platform (such as Twitter where people can and do use pseudonyms), identity often cannot be escaped. Modeling can ferret out many characteristics in a probabilistic but highly reliable manner Kossinki (2013). Commercial databases which match computer IP to actual voter names for the overwhelming majority of voters in the United States (Campaign Grid, 2012) are now available—thus, political campaigns with resources can now link individual computers to actual users and their computers without the consent, or even the knowledge, of the person. Third, one characteristic of big data is that anonymity is very difficult to maintain—as computer scientists have shown repeatedly, large datasets of user information can often effectively be deanonymized because, given enough data points, most profiles end up reducing to individuals (for example, just date of birth, gender and zip code positively identifies almost 90 % of people in the United States). Overall, there are very few ways in which a citizen can engage a campaign without being fully embedded in her identity while what the campaign knows about her, or has tailored to her, remains opaque and unknown.

Discussion and Conclusion

The critical theory of the 20th century is littered with dystopias. George Orwell's 1984 introduced nightmares of an all-knowing state, and Aldous Huxley explored the terror hidden in a world where individualized seduction is delivered through subliminal conditioning and drugs (soma). Yet, the 20th century ended without such horrors and 21st century ushered in new digital technologies that brought about new opportunities for participation, organizing and collective

action by citizens. Social movements around the world, ranging from the Arab uprisings (Tufekci, 2012) to the Occupy movement in the United States (Gitlin, 2012), have made use of these new technologies to organize dissent against existing local, national and global power. Such effects are real and surely they are part of the story of the rise of the social Internet. It would be naïve, however, to think that the impacts of these technologies would remain solely disruptive, or only contain dynamics that favor challengers (Morozov, 2012).

History of most technologies shows that those with power find ways to harness the power of new technologies and turn it into a means to further their own power (Spar, 2003). From the telegraph to the radio, the initial period of disruption was followed by a period of consolidation in which the challengers were incorporated into transformed power structures, and disruption gave rise to entrenchment. There are reasons to think that the Internet's trajectory may have some differences—its rate of diffusion exceeds any historical parallel, so it has been more firmly established. However, there is little reason to think that it will escape all historical norms.

Antonio Gramsci warned that elite power perpetuates itself as much, if not more, through cultural hegemony than through force or coercion; Edward Bernays underlined the systematic, scientific nature of "engineering of consent" by the rulers. And Theodor Adorno identified the maintenance of elite power through collusion of mass marketing with politics as the creation of a world of "something provided for all so that none may escape." This may yet be the formulation of a more potent form of hegemonic entrenchment compared with the fear-driven surveillance of 1984, and yet emerging fully as a possibility only now. And this would occur in an environment of total information asymmetry (Kreiss, 2012) so that the ordinary citizen, whose privacy is upended by this data, and who is surveilled at every turn, is not even aware of either the amount of data, the specifics of the analysis, or the individuality of the targeting directed at her.

It's worth pondering if the "insurgent" campaign of Obama 2008 would be possible in 2012, where Obama's data advantage played the starring role. The Obama 2012 campaign reportedly spent $150 million on its data operations, 15% of its budget, (and totally outspent the Romney campaign in this area). Proprietary databases required for big data analytics are costly, not widely available as they are often controlled by private platforms, and the equipment and expertise required to effectively use the data is extensive. At a minimum, this environment adds to dynamics which favor incumbents who already have troves of data, and favors entrenched and moneyed candidates within parties, and the data-rich among the existing parties.

The trends are clear; the selling of politicians as if they were "products" is not only going become more expansive, it's going to get better, if more expensive. In this light, it is not a complete coincidence that the "chief data scientist" for the Obama 2012 campaign was previously employed by a supermarket to "maximize the efficiency of sales promotions." And while the data advantage is held, for the moment, by the Democrats, it will likely available to the highest bidder in future campaigns.

The methods of computational politics will, and already are, also used in other spheres such as marketing, corporate campaigns, lobbying and more. The six dynamics outlined in this paper—availability of data, the shift to individual targeting, the potential and *opacity* of modeling, the rise of behavioral science in the service of persuasion, dynamic experimentation, and the growth of new power brokers on the Internet who control the data and algorithms—will impact many aspects of life in the 21st century. More direct research, as well as critical and conceptual analysis, is crucial to increase both our understanding and awareness of the new information environment, as well as to consider policy implications and responses. Similar to campaign finance laws, it may be that information use in elections needs regulatory oversight

due to its impacts for power, campaigning, governance and privacy.

This is not literally a world in which "something is provided for all so that none may escape." The primary reason that opacity is built so integrally into this system of surveillance, big data, and computation, unlike in Bentham's Panopticon, is that we are citizens, not prisoners, and citizens are still capable of exercising agency and power, and resisting hegemony. Starting an empirically informed, critical discussion of data politics for the 21st century, by deepening our understanding, may be the first important step in asserting our agency with respect to the big data that is generated *by* us and *about* us, but is increasingly being used *at* us.

References

Adorno, T. W., & Horkheimer, M. (2002). Enlightment as Mass Deception. In *Dialectic of Enlightenment: Philosophical Fragments* (pp. 94–136). Stanford University Press.

Anderson, N. (2008). Universities launch elephantine 78 terabyte digital library. *Ars Technica*.

Barocas, S. (2012). The price of precision: Voter microtargeting and its potential harms to the democratic process. In *Proceedings of the workshop on Politics, elections and data*

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*(7415), 295–298. doi:10.1038/nature11421

Bernays, E. L. (1947). The Engineering of Consent. *The ANNALS of the American Academy of Political and Social Science*, *250*(1), 113–120. doi:10.1177/000271624725000116

Campaign Grid. (2012). Campaign Grid 2012. Retrieved from www.campaigngrid.com.

Carty, V. (2011). *Wired and mobilizing: social movements, new technology, and electoral politics*. New York: Routledge.

Constine, J. (2012). How Big Is Facebook's Data? 2.5 Billion Pieces Of Content And 500+ Terabytes Ingested Every Day | TechCrunch. *TechCrunch*.

Dewey, J. (1927). *The public and its problems*. H. Holt and Company.

Epstein, E., & Robertson, R. (2013). Manipulating Search Rankings Can Shift Voters' Preferences Substantially Without Their Awareness. APA, Washington, D.C., May 2013

Foucault, M. (1977). *Discipline & punish*. Random House of Canada.

Fraser, N. (1990). Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy. *Social Text*, (25/26), 56–80. doi:10.2307/466240

Habermas, J. (1991). *The structural transformation of the public sphere: An inquiry into a*

category of bourgeois society. MIT press.

Hillygus, D. S., & Shields, T. G. (2009). *The Persuadable Voter: Wedge Issues in Presidential Campaigns*. Princeton University Press.

Howard, P. N. (2005). *New media campaigns and the managed citizen*. Cambridge University.

Howard, P. N., & Kreiss, D. (2010). Political parties and voter privacy: Australia, Canada, the United Kingdom, and United States in comparative perspective. *First Monday*, *15*(12). doi:10.5210/fm.v15i12.2975

Howard, P. N., & Parks, M. R. (2012). Social Media and Political Change: Capacity, Constraint, and Consequence. *Journal of Communication*, *62*(2), 359–362. doi:10.1111/j.1460-2466.2012.01626.x

Gerber, A. S., & Green, D. P. (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review*, 653–663.

Gitlin, T. (2012). *Occupy Nation: The Roots, the Spirit, & the Promise of Occupy Wall Street*. It.

Green, D. P., Gerber, A. S., & Nickerson, D. W. (2003). Getting Out the Vote in Local Elections: Results from Six Door-to-Door Canvassing Experiments. *Journal of Politics*, *65*(4), 1083–1096. doi:10.1111/1468-2508.t01-1-00126

Issenberg, S. (2012). *The victory lab: the secret science of winning campaigns*. NY: Crown.

Karpf, D. (2012). *The MoveOn effect: The unexpected transformation of American political advocacy*. Oxford University Press.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1218772110

Kreiss, D. (2012a). Yes We Can (Profile You): A Brief Primer on Campaigns and Political Data.

Stanford Law Review Online, *64*, 70.

Kreiss, D. (2012b). *Taking Our Country Back: The Crafting of Networked Politics from Howard Dean to Barack Obama*. Oxford University Press.

Kreiss, D., & Howard, P. N. (2010). New Challenges to Political Privacy: Lessons from the First U.S. Presidential Race in the Web 2.0 Era. *Intl Journal of Communication*, *4*(0), 19.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., … Alstyne, M. V. (2009). Computational Social Science. *Science*, *323*(5915), 721–723. doi:10.1126/science.1167742

Lohr, S. (2012, February 11). Big Data's Impact in the World. *The New York Times*.

Lippmann, W. (1927). *The Phantom Public*. Transaction Publishers.

Nielsen, R. K. (2012). *Ground wars: Personalized communication in political campaigns*. Princeton University Press.

Plato. (2012). *Plato's Republic : a dialogue in 16 chapters*. (A. Badiou, Trans.). New York: Columbia University Press.

Rutenberg, A. N., Ashley Parker, Jim, & Zeleny, J. (2012, November 7). Obama Campaign Clawed Back After a Dismal Debate. *The New York Times*.

Scherer, M. (2012, November 7). Inside the Secret World of the Data Crunchers. *Time*.

Spar, D. L. (2003). *Ruling the Waves: From the Compass to the Internet, a History of Business and Politics along the Technological Frontier*. Mariner Books.

Shirky, C. (2008). *Here comes everybody: Power of organizing without organizations*. Penguin.

Siroker, D. (2010). How Obama Raised $60 Million by Running a Simple Experiment. *Optimizely Blog*.

Yagoda, B. (2013). The Comic Stylings of POTUS. The Chronicle of Higher Education.