Accepted Manuscript: F. M. Simon, 'We Power Democracy'. Exploring the Promises of the Political Data Analytics Industry. *The Information Society*

'We Power Democracy' Exploring the Promises of the Political Data Analytics Industry

Felix M. Simon*

European Journalism Observatory, University of Oxford

Abstract

Data-driven campaigning has moved into focus in recent years. Yet, we still have a limited understanding of the emerging industry of political data analytics companies: how they envision data analytics and voter targeting, their role in electoral processes and which promises they make to their clients. This article focuses upon the way in which these topics are conceived of in the marketing rhetoric of the political data analytics industry. Drawing on a sample of 19 political data analytics companies it systematically explores in which ways data analytics is envisioned and marketed as a powerful tool in electoral processes, exposing a fundamental disconnect between scholarly discourse on the one hand–often critical of the claims of these companies about the efficacy of their methods–and a highly functionary data imaginary on the other hand, actively fostered by the political data-analytics industry and the media.

Keywords political campaigns; big data; data analytics; social media; political communication

Introduction

In the wake of the Brexit referendum and the 2016 U.S. presidential elections we have seen a growing interest in the role of data analytics, digital-campaigning and voter targeting in election campaigns. Both public commentary in the media and academic research have focused in depth on the role played by data analytic companies – most

*F. M. Simon, 2019. 'We Power Democracy': Exploring the Promises of the Political Data Analytics Industry. The Information Society: doi:10.1080/01972243.2019.1582570. This post-print last updated Feb 12, 2019. Please consult the final version of record for page numbers and references.

notably former UK-based firm *Cambridge Analytica* –, and attempts to micro-target and persuade voters in the run-up to both events (Baldwin-Philippi 2017; Bodó et al. 2017; Chester and Montgomery 2017; Illing 2017; Kreiss and McGregor 2017; Kuchler 2016; Weaver 2018). While much has been written by journalists (see, for instance, Cadwalladr 2017; Cadwalladr 2018; Grassegger and Krogerus 2017; Halpern 2017) about the assumed effectiveness of these efforts, our understanding of how this industry itself conceives of data analytics for political purposes remains patchy.

This article focuses upon the ways in which data, data-driven campaigning and voter targeting are conceived of in the marketing rhetoric of what I term the *political data-analytics industry*. I argue that to understand the spread of data analytics for political purposes as well as the recent controversy around this topic, it is pivotal to take a closer look at the promises the respective companies make around the data they gather as well as its analysis and deployment for political ends. Similarly, I ask how these firms conceive of the data itself and its seemingly magical potential in deciding elections in order to understand how this self-perception and description is mirrored in contemporaneous public discourse.

To this end, I focus on a sample of 19 political data analytics companies as typical representatives of an increasingly important industry in electoral politics around the world. Using this sample, this article scrutinises the visions and promises the political data analytics industry creates around data, its availability, analysis, deployment for (and effectiveness in achieving) certain political ends. It further tries to understand how the self-presentation of this industry has shaped its perception in current public discourse.

Following a brief review of relevant literature, I provide a description of the methodology. In a second step, I present the findings drawn from the material and discuss them in the context of present research on the effectiveness of data-driven campaigning, voter targeting and the use of big data in politics.

Big Data, Data-Driven Campaigning and Electoral Politics

As Anstead (2017) notes, recent years have seen a growing interest in the role of data and data-driven campaigning in election campaigns, particularly in the United States, where to this date data-driven campaigning has been used most widely. This interest has intensified and reached a wider audience throughout the 2008 and 2012 U.S. presidential elections and arguably peaked following the Brexit referendum and the 2016 U.S. presidential election. Here, data-driven campaigning in general and micro-targeting of voters in particular was rumored to have tipped the scales for the Leave campaign and Republican candidate Donald Trump, respectively.

Employing the term 'computational politics', Tufekci (2014) defines digital data-driven campaigning as 'applying computational methods to large datasets derived from online and off-line data-sources for conducting outreach, persuasion and mobilization' (2). Tufekci stakes out a rough outline of the political data analytics industry which is, according to her, 'informed by behavioral sciences and refined using experimental approaches, including online experiments' in order to profile people and 'to develop methods of persuasion and mobilization'. More specifically, data-driven campaigning involves two key features: a) *targeting*, that is 'deciding which messages go to what

potential voters at what time during the campaign' and b) *testing*, the act of 'empirically measuring how well messages perform against one another and using that information to drive content production and further targeting' (Baldwin-Philippi 2017, 2). Tufekci contends that this set of practices is crucially shaped by the existence of big data and 'accompanying analytic tools', many of which were first developed, tested, and refined by the commercial sector, especially in marketing and advertising (Vaidhyanathan 2018).

As Baldwin-Philippi (2017), O'Neil (2016) and Hersh (2015) note, academics and practitioners have tried to harness public voting records, census data, credit card purchases and other publicly available data since the 1990s to improve voter targeting. Nielsen's (2012) ethnographic study of a congressional 'ground war' in 2008 spotlighted the role of technical infrastructures, large databases and high-tech targeting technologies in bringing about 'personalized political communication' to engage directly and more efficiently with the electorate. A decade later, data-driven campaigning is institutionalized as a significant part of political campaigns around the world (Anstead 2017; Kruschinski and Haller 2017), especially in the U.S. Increasingly, these methods are also employed by stealth actors in an attempt to swing elections. As recent research by Kim et al. (2018) has demonstrated, stealth actors have attempted to target and manipulate voters in the U.S., raising questions about the role of foreign and other non-state actors trying to influence elections by harnessing the techniques of data-driven campaigning.

Technology companies are now key actors in political campaigns, providing a wide range of digital marketing tools and techniques for political use (Chester and Montgomery 2017). Kreiss and McGregor (2017) observe that leading technology firms such as Facebook, Twitter, Microsoft, and Google have been actively assisting campaigning efforts during the 2016 U.S. presidential election campaign. By developing 'organizational structures [...] that accord with the partisan nature of American politics', these firms partake in creating and shaping campaign communication through 'close collaboration with political staffers' (1). While the role of major tech companies is one important aspect, Chester and Montgomery (2017) highlight that electoral politics have also become 'fully integrated into a growing, global commercial digital media and marketing ecosystem' with a wealth of smaller, 'specialized firms' offering their services to campaigns on both sides of the political aisle (2) – what I term the political data analytics industry.

What unites all these authors is the shared conviction that (big) data, its analysis and its deployment to target and persuade voters is permanently shaping political campaigns and likely to only intensify over the coming years – not least thanks to technological progress in terms of computing power, data gathering and analysis capabilities as well as the increasing availability of fine-grained personal data through social media (see Tufekci 2014; Nickerson and Rogers 2014). At the same time, a recurring element in much of the literature is the observation of a methodological shift over time, away from broad area-based targeting towards a more individualized approach (Anstead 2017, 5), ideally addressing each member of the populace with a tailor-made message in accordance with their characteristics (Nickerson and Rogers 2014): what is now more commonly known as *micro-targeting.* A final observation shared by these sources is that data and data analysis are treated by campaigns as a cost-saving measure. By facilitating a more efficient allocation of scarce resources (Anstead 2017, Vaidhyanathan 2018) campaigns ultimately seek to increase the marginal product in terms of voters persuaded or voters led to the ballot.

Tufekci and others (see for instance O'Neil 2016) have noted that the rise of data-driven campaigning is closely tied to the rise of 'big data' and big data analytics. Within the scope of this article, it is not possible to go into the already extensive discussion of the nature and definition of 'big data' (see for this Schroeder 2018a). Nevertheless, the concept merits a brief definition as it is useful in contextualizing the promises of the political data analytics industry and grounding them in reality.

Briefly, the term is understood to be data that is 'unprecedented in scale and scope in relation to a given phenomenon'. Rather than only referring to fixed datasets, big data can be seen as large-volume streams of data from a multitude of sources accumulating at high velocity (Meyer and Schroeder 2015, 144) – for instance, consumer purchases in online stores, data generated on social media or through (mobile) devices and increasingly the data generated by the 'Internet of Things', often gathered without the knowledge or explicit consent of users (see Turow 2017). boyd and Crawford (2012) also make out the capacity 'to search, aggregate, and cross-reference large data sets', further noting that big data, asides from its technical aspects, entails a cultural ideology, circulating around the notion that the accumulation and analysis of large datasets yields a new form of a more truthful and objective knowledge of the world (663). While scholars have cautioned from the beginning that big data cannot work magic, the exceptionalism inherent in the big data paradigm often purports that it has near-magical properties in comparison to other technology. In a political context, this magic has repeatedly been described as potentially having brought about the outcome of recent political campaigns (Krogerus and Grasegger 2017), and to have 'sway[ed] elections in ways that people can't even see, [and] don't even realise is happening to them' (Cadwalladr 2017). A key term in this context is 'managed democracy': With the help of very fine-grained personal data collected at a grand scale, voters are purportedly being manipulated and covertly influenced in a way to either motivate them to vote for a certain cause or candidate or to stay away from the ballot box – all in order to swing an electoral process in a direction desired by the group or individual 'pulling the strings'. The political data analytics industry–as we will see–both claims and is claimed to play a decisive role in these processes.

With this background in mind and in order to untangle some of the claims and perceptions formed around the political data analytics industry, this article addresses two main research questions:

I. *What promises are made by the companies involved in political data analytics with regard to the data they gather as well as their analysis and deployment for political ends?*
II. *How does the political data analytics industry conceive of the data and data analytics and its seemingly magical potential in deciding elections?*

Taken together, these research questions will provide a framing for this analysis of the political data analytics industry, and help us develop a more general hypothesis of the visions and promises they create around data, its availability, analysis, deployment for (and effectiveness in achieving) political ends and how this industry is perceived in current public discourse.

Identifying the Political Data Analytics Industry

To understand how data, data-analytics, data-driven campaigning and voter targeting are conceived of in the marketing rhetoric of the political data analytics industry it was necessary to first identify a large enough sample of data analytic companies working on data-driven campaigning whose activities and online presences could then be examined in detail. Following the approach by Beer (2018a), the sample for this article was created between the 10 January and 15 February 2018 by first using two different search engines (Google and DuckDuckGo) and a combination of seven different search terms. These search terms were: (1) data-driven campaigning; (2) data analytics; (3) data analytics companies; (4) election management; (5) voter targeting and (6) micro-targeting. Based on the existing academic and journalistic coverage of the topic, these search terms seemed to be the most appropriate to locate the respective organisations as well as articles and similar sources which focused on the same.

A second, more straightforward approach was to enter the term 'Companies similar to Cambridge Analytica' (7) into both search engines and peruse the results. Following both the Brexit referendum as well as the U.S. presidential election in 2016, *Cambridge Analytica* gained renown as having allegedly played a significant role in both. Consequently, the company–which in the meantime has filed for insolvency and closed operations after revelations by British newspaper *Observer* that it had illegally gathered Facebook user data–could be assumed to be the most publicly known organisation in this field and a likely point of reference in articles and lists on the topic of digital campaigning and data analysis for political purposes.

The sample was then created by manually going through the first 20 entries for each of the combinations between search term 1 and any one of the other search terms 2, 3, 4, 5 and 6, as well as by going through the first 20 entries for search term 7. The combination of search terms 1 and 3 and 1, 5 and 6 as well as 7 generated the most diverse and useful lists of the type company of interest for this article. The top 20 hits for these search terms included several articles (e.g. Kuchler 2016; Mitchell 2012) which provided overviews of digital campaigning companies. They, too, were used to create the final sample. By visiting the website of every single company identified through the search and by triangulating the findings with other sources (e.g. media coverage and Wikipedia entries) I aimed to only include those companies in the final list which had reliably worked on political campaigns in the past (as evidenced by references to previous clients, campaigns and additional research).

Once the final sample had been established (see the complete list in the appendix) I then studied the entire web presences (e.g. mission statements, company descriptions, listings of services, testimonials, etc.) and marketing materials (e.g. press releases, case studies, promotion videos) on the websites of the respective companies. In line with the research questions and the aims of this article, I focused specifically on those materials that were related to politics and electoral campaigns (some companies offer their services both to campaigns and the commercial sector). I first looked at the range of different services, solutions and products that the companies advertised to potential clients on their websites. The motivation for this was to understand a) which specific problems the firm claimed to address and which solutions it proposed, b) which opportunities it promised to create and c) which type of data and data analytics it used (or promised to use) in addressing these problems or creating these opportunities. Second, I scrutinized – if available – the case

studies and testimonials on the websites of the companies in the sample, to understand the context in which (and to what ends) the advertised approaches have been used. Third, and finally, I directed my attention on the self-presentation of the respective companies. How did they present themselves to potential clients (e.g. information about costs, contact details)? How transparent or opaque were they about the services and promises they offered? These questions served as guidelines to analyse the source material in a meaningful way and to find answers to the research questions posed in the beginning.

Findings

For a start, a few general remarks about the companies studied are required. The political data analytics industry as studied in this context (owing to the search terms) is a predominantly U.S.-American phenomenon. Out of the 19 companies in the sample, 17 were U.S. based, with the exception of *AggregateIQ* (Canada) and *Cambridge Analytica* (United Kingdom), something that is reflected in the existing media coverage as well as in the scholarly attention the topic has received in the past. Among some of the U.S.-based companies, a split between those working for progressive causes and campaigns (7 in total) and companies working for conservative causes and campaigns (7 in total) emerged (the remaining three had no clear affiliations).

Further worth mentioning is the positioning of the political data analytics industry. Contrary to what the term suggests, many companies can be located at the nexus between the political and commercial sphere. This highlights not only their historical genesis out of the marketing and advertising industry as well as political consulting (Harding 2008, Vaidhyanathan 2018) but also an apparent interest on part of these companies to diversify their income streams by offering their services both to the realm of business and politics. Out of 19 companies in the sample, 11 presented themselves as being solely focused on political campaigns and initiatives, while 8 offered their services both to commercial and political clients.

A recurring theme across the companies in the sample was a concentrated effort to present their methods as grounded in science and to conjure up an image of their work as being conducted with scientific rigour. Many companies speak of 'scientifically grounded analytics' (Appendix, ref 5), a 'scientific focus' (Appendix, ref 7, 19) and showcase their human capital (for instance, by listing their degrees in STEM, statistics or political science and the often prestigious universities they attended). One of the companies states that its efforts are 'headed by PhD Data Scientists' (Appendix, ref 9) while another makes reference to their 'multitalented workforce of researchers, data scientists, behavioral psychologists [...]', among them 'PhDs from the world's leading universities [...]' (Appendix, ref 1) While the work carried out by these firms may not be scientific – at least not in a sense of following scientific codes of conduct, adhering to scientific ethical standards or in enforcing scientific rigor – the prominent display of 'scientific' attributes may be seen as a form of virtue signaling, reminding potential clients and in a wider sense the public and the media about the seriousness and trustworthiness of their business. It also acts as a marker of distinction from other companies in the same field. Or, as one firm puts it: 'We don't just roll the dice based on small focus groups or best practices – we use science' (Appendix, ref 3).

Another key promise is that of increased efficiency and more value for money. Services are pitched as a cost-saving measures which help to increase the marginal product

of campaigns in terms of voters persuaded or voters led to the ballot. Their self-presentation in this regard is straightforward: 'Our proprietary formula identifies the programs that provide the best bang for your buck – reaching ideal targets and saving money', claims one (Appendix, ref 3) while others praise themselves as a 'cost effective option' (Appendix, ref 4). Targeting voters online, it is claimed, 'will save campaigns money and truly impact the way candidates and causes are viewed' (Appendix, ref 13). Implicitly acknowledging the stellar costs of political advertising, a special emphasis is put on spending funds where they will likely yield the best results. In this vein, one company in the sample urged potential clients to 'stop wasting time and money marketing to the wrong people' (Appendix, ref 17), a notion also found at other firms where they claim that they are able to 'deliver relevant messages to the most receptive online audience while reducing inefficient ad spend' (Appendix, ref 15).

A final remark pertains to the wide margin in available information. A bare minimum of information was provided by Canadian data analytics firm *AggregateIQ* whose website only gave away a minuscule amount of information about its practices. The opposite end of the spectrum was inhabited by companies such as *Cambridge Analytica* and *HaystaqDNA* who, at the time of writing, provided comparably much (though superficial) detail about their operations. Most firms staked out a middle ground, trying to advertise their solutions without giving away too much detail about their techniques. Having addressed some general observations, the rest of the article explores the presences of the respective companies in greater detail.

'Superior Data' – Data Collection and Data Quality

One overriding promise which emerged from the studied materials is the superior *comprehensiveness*, *granularity* and *liveness* of the data which these companies gather and use in their analyses. The scale of the datasets, the type and number of collected attributes and in some cases the sources of the data are frequently displayed on websites and in marketing materials and used to signal the firm's advantage over others in the same field as well as to cement the idea that to win it is pivotal to rely on their services.

We are told that companies collect 'every possible signal from each person' (Appendix, ref 3) and provide 'the most comprehensive voter data, consumer files, and donor files anywhere' (Appendix, ref 10) in order to deliver an 'unmatched understanding of your electorate' which will, for instance, help to 'pinpoint the voters who will turn the tide in your favor' (Appendix, ref 1). This form of big data purportedly provides the 'fullest possible picture' and a 'rich, holistic view of voter behavior' (Appendix, ref 1).

The promises around the collected data are very much a game of numbers. Different firms boast that they have a 'database of 220 million Americans' (Appendix, ref 3), '5,000 data points on over 230 million American voters' (Appendix, ref 1) or 'data on over 240 million unique voting-age individuals across all 50 states' (Appendix, ref 9), although many choose not to give any specifics, in particular as to origin of the data. The same can be said for the types of collected attributes. While some companies do not disclose any information at all (see, for instance, Appendix, ref 2, 16), others state that they collect 'data on attitudes, behaviour, personality' (Appendix, ref 7) or 'variables like demographic information and buying behaviour' (Appendix, ref 5). Only a few firms go into greater detail and claim that they collect data across 'hundreds of fields, including household

attributes, purchasing and investment profiles, donation behavior, occupational information, recreational interests, and engagement with civic and community groups' (Appendix, ref 9, 14).

A thread that runs through the sample is an emphasis on the *liveness* of the data. Very much in line with Meyer and Schroeder's (2015) assertion that a key definition of big data is its accumulation at high velocity, the companies pride themselves on having access to 'a constant stream of data', 'the most up-to-date voter and consumer files possible' and that they use 'real time data' (Appendix, ref 7, 4, 3). 'Dynamically updated' and 'continuously updated' (Apprendix, ref 4, 9) are prominent terms to describe the nature of the datasets at disposal of the companies and underscores a general, underlying message that only with data that is continuously amassed at a large scale the client can hope to get an edge over her competitors. As Beer (2018a) has noted, data analytics companies often prize their data as allowing for 'a representation of the world as it unfolds – rather than being a reflective process of looking back' (465), an idea that is also evident in the self-description of the political data analytics industry.

A third point is the *range of data sources* to which a company has access to. These, too, are used as markers of distinction, setting apart one firm from another in the competition for clients. A majority of companies use existing commercially available aggregated voter file data for the U.S. as a whole or create their own national voter file by aggregating data from publicly available government records of who is registered to vote and who cast ballots in past elections. However, as the firms in the sample readily admit, these sources are usually enhanced with other commercial and proprietary data sets, the company's own data or data supplied by the client. One company describes their sources as follows: 'A reliable foundation of 185 million registered voters – a unified national voter file – collected from Election Officials' which is then supplemented by 'commercial data [which] adds another 55 million unregistered individuals and enhances the entire database'. They further claim to 'integrate data from the Census, specialty data, and media market geographies' (Appendix, ref 9). In a similar vein, some of the other companies studied claim to enhance the national voter file with 'five national telephone source files, current U.S. Census Data, election return data from every county nationwide [...] and comprehensive lifestyle data and modeling scores on relevant issues' and claim to be in possession of 'non-registered voter data and email addresses' (Appendix, ref 14). Another company highlights their access to 'anonymous behavior from mobile and desktop browsing activity of 90% of U.S. adult web users' as well as 'conduct[ing] 200,000+ online interviews' (Appendix, ref 15). Yet another common addition is survey data. The companies frequently mentioned that they have access to, for instance, 'the nation's largest proprietary voter survey data' (Appendix, ref 15) or other state-of-the-art survey data which is gathered to complete the picture. One uniting feature is that no company gives – apart from the national voter file – many specifics about where they get the data from and how detailed these datasets really are.

Taken together, these 19 examples that companies promise data that is unrivalled in its comprehensiveness and granularity, allowing campaigns to identify and target individual voters, as well as data that is always up-to-date, satisfying the demand for real time insights and action. The imaginary promised by these firms is that of data and analytics which enables campaigns to tap into every individual voter's mind and heart, at every point in time, giving campaigns 'granular insight into [their] audience' (Appendix, ref 8).

Opaque Analysis

When it comes to the actual analysis of the data, the companies cloud themselves in mystery. The description of their methods never goes beyond a superficial level, revealing very little about the underlying mechanics at work. Information about their practices, if it exists at all (see, for instance, Appendix, ref 2) is often limited to a few, vague sentences. Nevertheless, we can make a few general observations about the methods which they claim to apply.

For one, modelling – the process of creating, testing and validating a model based on sample data to best predict the probability of an outcome – plays a major role. We are told that companies use 'predictive modelling' as well as 'look-alike modelling, survey-based modelling' (Appendix, ref 3, 5, 18, 19) based on their own data or data provided by the client to find 'the right individuals' or 'likeminded individuals' – that is, individuals which match the clients interests or are potentially most susceptible to their messages. Frequently mentioned are also 'advanced statistical models, [which allow to] map groups onto the entire population at the individual level', 'applied machine learning' (Appendix, ref 5), 'statistical and machine learning algorithms' (Appendix, ref 15) which enable the client to segment the electorate into distinct audiences and 'artificial intelligence that takes into account the behavioral conditioning of each individual to create informed forecasts of future behaviour' (Appendix, ref 1).

In summary, it can be said that the companies do not seem to have a strong interest in disclosing any particulars about how exactly data are processed and analysed. Rather than providing any details, we are told about 'machine learning', 'statistical models' and 'artificial intelligence' – keywords which can potentially mean anything and nothing at the same time and are often imbued and associated with a vague notion of innovation, modernity and progress. From an entrepreneurial perspective, it certainly makes sense not to give too much away – after all, these companies are in direct competition with each other. However, shrouding oneself in mysteries by only giving away little information and letting the imagination 'run wild' also contributes to a situation in which a certain omnipotence is quickly ascribed to the methods of these companies.

Deployment and Promises – 'We don't just talk about big data. We make big wins.'

While the political data analytics industry is tight-lipped about the acquisition and analysis of user data, it is more open about the actual deployment – the ends to which data is gathered and analysed. It is in these statements, that the companies create a data analytics imaginary that is by now etched into most people's minds and has significantly influenced the discourse around digital campaigning and the perceived influence and effectiveness of such efforts in recent elections.

What is striking are the bold claims made by most companies, *claims which are essentially interchangeable and often promise clients quasi-magical power by harnessing 'big data'.* 'Test Everything. Change Behaviour. Drive Results', says one, while another promises to 'supercharge social movements and political campaigns' (Appendix, ref 3 and 7). We are told that a company is the leading expert in 'finding, understanding, and persuading people to vote a certain way', and that they will find 'our' voters 'and move

them to action' (Appendix, ref 1). Another company highlights that they 'don't just talk about big data. We make big wins' (Appendix, ref 11), while a competitor claims that it 'will equip you [the client] with the data and insights necessary to drive your voters to the polls and win your campaign' (Appendix, ref 1). Among the most militaristic statements was an offer to 'start building [a] digital army' which could be used to 'convince undecided voters, motivate [the] base and increase turnout' (Appendix, ref 6). In the self-presentation of the political data-analytics industry, the political realm is conceived of as a battlefield where data analytics enable campaigns to 'gain advantage over your opponents' and 'pinpoint the voter who will turn the tide in your favor' (Appendix, ref 1).

A uniting trend is that *the companies position themselves as the pivotal element when it comes to raising funds, encouraging supporters and achieving wins at the ballot box by persuading voters.* In the words of one company, they offer 'the best data and technology to reach every voter, on every device, everywhere' (Appendix, ref 13). Slogans such as 'Your revolution starts here!', 'You need superior data.', 'Better Data. Better Decisions', 'We play a pivotal role in winning presidential elections as well as congressional and state elections.', 'To Move the Needle, You Have to Find It – Making Big Data and Analytics work for you!', and 'We don't just talk about big data. We make big wins.' (Appendix, ref 12, 14, 1, 17, 11) are reminders to potential clients of the importance of the political data analytics industry and big data analysis for political ends. Or in the words of yet another company: 'We help you cut through the clutter, make an instant impact on voters and start moving the needle' (Appendix, ref 6). The overriding message is: Without our support and services, your efforts will stall.

An important element of these services is the promise of *tailoring messages to the right individuals.* Companies in the sample promise to 'deliver an omni-channel, omni-message approach to tailor exactly the right outreach to the right person, and to identify the optimal distribution of resources across channels' (Appendix, ref 3), to 'target precisely the voters needed to win' and to help the clients 'identify the best messages and channels to engage your supporters' (Appendix, ref 13 and 5). The focus here is distinctively on these companies' purported ability to differentiate between individual voters, not just audience segments, and getting the 'right' – that is a tailor-made – message in front of each and every individual in order to persuade them or move them to action.

In this spirit, one company promises to 'identify and score every individual voter, determining the persuasive messages that will make the most impact' and boasts of its ability in 'persuading and mobilizing the right voters with the right messages' (Appendix, ref 11). Other firms, too, make such claims about their abilities and tell prospective clients that they can 'target the right message to the right voter' (Appendix, ref 6), 'find who you can persuade and activate [to] identify your most likely supporters' (Appendix, ref 8) and that they will 'dig deep into the data to understand your audience and its journey, test everything to find winning strategies, tailor outreach to each person, and bring these processes to scale through analytics technology and machine learning' (Appendix, ref 3). Another company frames all this slightly differently and explains that their 'technology enables the mass-individualization of campaigns – recognising that each individual may have different needs and tastes that require a different approach of persuasion', noting that they will 'amplify your core messages and help you build more engaged relationships with the right people at the right time' (Appendix, ref 7).

A final claim made repeatedly in this context is that big data analytics allow these companies not only to identify the 'right' individual and how to target them (or, to put it

differently, with what) but *also where to target them*, hence, which mediums individuals frequently use. Campaigning, so it is claimed, is made much easier and more efficient 'by finding the most likely individual supporters, and engaging them through the channels they are most likely to consume' (Appendix, ref 3). Many companies argue that they can 'pinpoint the best marketing channels' (Appendix, ref 8) and 'leverage many platforms to deliver messaging across desktop, mobile, tablet and connected TV devices', using channels which include 'display, video, Facebook, Twitter, native, audio, interactive and search' (Appendix, ref 1).

Discussion

In the eyes of the public, data-driven campaigning and micro-targeting are currently perceived mainly through a dystopian lens. Concerns about the fairness of electoral campaigns and the future of democracy are recurring talking points in this discourse. Recent years have seen the persistent rise of a narrative in which especially political micro-targeting is described as bringing about a future where attitudes and beliefs can be manufactured at the individual level (Kreiss 2017), what Howard (2006) calls 'managed citizenship'. As Baldwin-Philippi (2017) contends, many of these dystopian predictions envisage a world where citizens are increasingly caught in 'filter bubbles', succumb to 'echo chambers' and are generally no longer able to inform themselves accurately and independently. In this particular narrative, pushed through extensive and at times sensationalist media coverage, data analytics companies appear as the ultimate villain of democracy – uncanny entities with advanced technological tools at their disposal; 'Bondesque' shadow organisations, ready to thwart democratic electoral processes at their master's orders.

Given the findings of this research, it becomes clear why this picture could have developed in public discourse. Predominantly presenting themselves as decisive actors in modern-day political communication, the data analytics companies, as seen in the sample, boast of their abilities in influencing voters and winning election campaigns through big data analytics – without delivering substantive proof. They allege to have unrivalled access to different types of data on individuals which – it is claimed – are only superior to the data of their competitors, but are also the pivotal element in election campaigns, bringing about electoral success with a simultaneous increase in efficiency and a reduction of costs. However, the exact access to and the exact origin of the data is – in most cases – never described in any greater detail, giving them a mystical aura. The same applies to the specific methods of analysis and the final use of the data in targeting and persuading voters, both of which are predominantly described superficially and often in a way that gives free rein to the imagination of outsiders.

In addition, the close reading of the marketing materials has not only identified the promises of the political data analytics industry – that is, game-changing results obtained through state-of-the-art big data analysis and messaging – but also shed light on their conception of data and data analytics as well as its magical potential. Through their self-presentation they create what Beer (2018a; 2018b) has termed a data imaginary in which powerful promises are projected onto data and data analytics. These analytics are said to be revealing with prophetic capabilities, enabling us to observe unseen truths at any time, everywhere. The specific imaginary created by the industry is one of data that is unrivalled

in its comprehensiveness and granularity and always up-to-date, thus allowing campaigns to identify and target every individual voter and ultimately to persuade them to act, donate or vote in line with the clients respective interests. It is also closely aligned with the idea of the increasing 'rationalization' of audiences, as described by Napoli (2010). In Napoli's argumentation, new technologies have enabled an increasingly powerful and systematic measurement of audiences, allowing for greater control in terms of which audiences are targeted and how messages can be tailored to their interests. Whether knowingly or unknowingly, the companies in the sample take up this notion and position themselves as the manifestation of this prophecy. Just as the forefathers of modern political consulting sold themselves as key factors in winning elections (Harding 2009), so do their successor on the political playing field of the 21st century.

It is then perhaps not particularly surprising that the imaginary of the omnipotent data analytics firm should have found its way into wider public discourse, especially following the revelations around *Cambridge Analytica*. Much of the journalistic coverage on the topic, despite the occasional contrarian take, has largely uncritically accepted these companies' own claims about their effectiveness, passing them off as either a pivotal element in recent elections or a serious threat to modern-day democratic processes (BBC Two 2017; Cadwalladr 2017; Cadwalladr 2018; Grassegger and Krogerus 2017; Susskind 2018). A public talk at the 2016 Concordia Annual Summit of former *Cambridge Analytica* CEO Alexander Nix where he boasted of his company's abilities was widely shared and reported upon in the media, often without questioning the veracity of his claims.

Yet, the reason why this imaginary has become so pervasive might lie not only in the overt self-promotion but also in the paucity of more detailed information about these firms and their practices. As this analysis has demonstrated, actual in-depth description of the gathering, analysis and deployment of user data for political ends rarely extends beyond buzzwords and superficial descriptions which disclose little about the methods at work. The marketing materials of these companies reveal next to nothing about success rates and instead, unsurprisingly, often rely on case studies where causal links are made between a firm's work and an electoral success story.

Unsurprisingly, these companies have a vested interested in upholding this narrative. Especially in the US, political data analytics firms operate in a hyper-competitive, yet also very lucrative market. According to Williams and Gulati (2017) Hillary Clinton and Donald Trump alone spent nearly $104 million between them on digital advertising during the 2016 election (7). Firms are in fierce competition with each other and can be expected to have a great interest in emphasizing their own significance and effectiveness in order to attract new or retain existing clients. The fact, that many political data analytics firms diversify their revenue streams by pitching their work to commercial clients, thus tapping into an even larger market, lends further credibility to this argument.

Yet, while claims about the magical potential of data analytics in a political context abound in public discourse, they are thrown into sharp relief by contemporaneous research on data-driven campaigning and micro-targeting which, as Anstead (2017) notes, 'has provided an important corrective to hyperbolic popular discussion.' To begin with, it is worth pointing out that the notion of omnipotence through data-driven campaigning has a long history and ties into long-standing fears of mass manipulation by new media and

technologies (Tworek 2017). In a political communications context, Anstead (2017) highlights that similar arguments to those found in current discourse were made in the past about 'billboard poster advertising, qualitative research and spin doctors' (6). To put it differently: The fear – kindled through the self-presentation of the political data analytics industry and its representation in the media – that data-driven campaigning will spell an end to democratic processes is hardly new.

While history helps us to see the claims made around these companies in in a new light, the available research throws further doubt on the magical potential of big data analytics for politics. As Hersh's (2015) seminal study on data-driven campaigning has shown, campaign practitioners find it very hard to persuade voters. Other political scientists, too, have emphasised the difficulty of influencing people and changing their political opinions with micro-targeting in particular (Hersh and Schaffner 2013) and advertising campaigns in general (Henderson and Theodoridis 2018; Kalla and Broockman 2017). As for the magical capabilities of *Cambridge Analytica* (ref 1) and its use of 'psychographics' (micro-targeting based on psychological characteristics), several authors have presented theoretical, experimental, and case study evidence that the claims were overstated and the method ultimately ineffective (Azucar, Marengo, and Settann 2018; Furnham and Fenton-O'Creevy 2018; Karpf 2017; Simon 2018).

Instead, how individuals cast their votes depends on a complex interplay of different factors – for example, their views on the economy or their educational background. Various studies have suggested that micro-targeting has comparatively little influence in this context (Broockman and Green 2014; Hersh and Schaffner 2013). For the U.S., Kreiss (2017) finds that partisanship already shapes much of how people vote. Others have suggested for U.S. presidential races that voters already know so much about the major candidates that 'ads offer little that is new to induce individuals to change their feelings and predisposition to vote for (or against) a candidate' (Fowler, Franz, and Ridout 2015; Williams and Gulati 2017). For the recent U.S. presidential election, various scholars have emphasised the role of social media, especially Twitter, in helping Donald Trump to get the attention of 'gatekeepers' in the mainstream media and thus being able to dominate the (television) news agenda (Baldwin-Philippi 2017; Cowls and Schroeder 2018; Schroeder 2018b) – a more decisive element when it comes to shaping voters perceptions and intentions, with cable news in particular being a significant driver of partisan divisions and vote shares (Martin and Yurukoglu 2017). According to Boxell, Gentzkow, and Shapiro (2018) Trump also performed worse than previous Republican candidates among internet users and people who got campaign news online, outperforming his predecessors only among the demographic groups least likely to be online. The authors thus suggest that the internet and micro-targeted advertising on social media was likely not a source of advantage to Trump. Finally, Neuman (2016) has cautioned about the assumed persuasive impact of advertising. Stressing the fundamentally polysemic nature of communication, he argues that any 'media message, intended to be persuasive or otherwise, is not likely to stimulate a singular response, but rather a distribution of responses across a population of those who have encountered the message' (81). Instead of the gullible and isolated pawns that are easily being mislead, Neuman points out, that most individuals are surprisingly good at ignoring advertising and propaganda (147) – an observation which makes the the claims of the political data analytics industry all the more questionable.

Conclusion

This article has tried to offer the most comprehensive and systematic analysis of the rhetoric of the political data analytics industry to-date. Unquestionably, political consulting firms are designed to strategically sell their wares so the fact that they are marketing themselves as described here should not surprise us. Yet, what this research adds to and demonstrates is a fundamental disconnect between scholarly discourse on the one hand–often critical of the claims of these companies about the efficacy of their methods–and a powerful data imaginary (Beer 2018a) on the other hand, actively fostered by the political data-analytics industry and spun further in the media. Knowing more about the self-presentation of these companies then helps us to comprehend why they could turn into such a menace in the public imagination in the first place and why their claims are often taken at face-value.

What this paper has shown is that these data analytics companies advertise themselves as using the most powerful scientific computational methods to process unprecedented, and previously unobtainable data, ascribing seemingly magical powers to both. But of course they cannot reveal what these magical powers are (there is a parallel here with digital media companies like Google that cannot reveal the basis of their search algorithm without losing their competitive advantage or opening themselves to 'gaming' their methods). Had it not been for the electoral surprises of Brexit and Trump, which were not just surprises but have fundamentally changed the course of British and American politics, these mysterious techniques might have gone unremarked. Because of these surprises, however, the public and the media have taken a deep interest in uncovering what will remain – in respect to the techniques discussed here – mysterious.

This can be put differently: in consumer marketing, where the effect of using these techniques has less critical consequences – affecting consumer choices – the techniques used can to some extent be revealed (see Turow 2017), though here too, some techniques cannot be disclosed because of competitiveness concerns. But in political campaigns, the mystery is unlikely to be cleared up (except in relation to illegal tactics or illegal data uses, which may require regulation but which are outside the scope of this paper) because politics is more critical to societal well-being. The political campaigns that avail themselves of these analytics services may to some extent be allowed to share in the mysteries of these techniques, but even here, the companies offering these services will need to retain some control over proprietary techniques and data vis-à-vis campaigns – lest they lose what makes their services valuable and profitable.

Magical powers will thus remain mysteries – and again, they now have the public and the media curious. With both the analytics companies and political campaigns aware that using these powers will face increasing scrutiny in light of the subsequent controversies (again, these are beyond the scope here) – this is the stalemated situation in which we now find ourselves with politics and data analytics. This situation is likely to remain so for the foreseeable future since it is unlikely that much light will be shone into the workings and efficacy of this magic.

Acknowledgments

I am indebted to Ralph Schroeder for taking the time to read and provide excellent feedback on various drafts of this paper. My thanks further extend to Michelle Disser and Erik Bucy for providing essential comments and support. Finally, many thanks to Harmeet Sawhney for his careful editing of the final version.

References

Anstead, Nick. 2017. 'Data-Driven Campaigning in the 2015 UK General Election'. *International Journal of Press/Politics* 22 (3): 294–313. https://doi.org/10.1177/1940161217706163.

Azucar, Danny, Davide Marengo, and Michele Settanni. 2018. 'Predicting the Big 5 Personality Traits from Digital Footprints on Social Media: A Meta-Analysis'. *Personality and Individual Differences* 124 (April): 150–59. https://doi.org/10.1016/j.paid.2017.12.018.

Baldwin-Philippi, Jessica. 2017. 'The Myths of Data-Driven Campaigning'. *Political Communication* 34 (4): 627–33. https://doi.org/10.1080/10584609.2017.1372999.

'BBC Two - Secrets Of Silicon Valley'. 2017. *BBC Two*. BBC. https://www.bbc.co.uk/programmes/b0916ghq.

Beer, David. 2018a. 'Envisioning the Power of Data Analytics'. *Information, Communication & Society* 21 (3): 465–79. https://doi.org/10.1080/1369118X.2017.1289232.

———. 2018b. 'The Data Imaginary: Six Reasons Why Data Analytics Have Become so Powerful'. *British Politics and Policy at LSE* (blog). 3 July 2018. http://blogs.lse.ac.uk/politicsandpolicy/data-imaginary/.

Bodó, Balázs, Helberger, Nathalie, & Vreese, Claes. H. de. 2017. Political micro-targeting: a Manchurian candidate or just a dark horse? *Internet Policy Review*, *6*(4). Retrieved from https://policyreview.info/articles/analysis/political-micro-targeting-manchurian-candidate-or-just-dark-horse

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. 2018. 'A Note on Internet Use and the 2016 U.S. Presidential Election Outcome'. *PLOS ONE* 13 (7): e0199571. https://doi.org/10.1371/journal.pone.0199571.

boyd, danah, and Kate Crawford. 2012. 'Critical Questions for Big Data'. *Information, Communication & Society* 15 (5): 662–79. https://doi.org/10.1080/1369118X.2012.678878.

Broockman, David E., and Donald P. Green. 2014. 'Do Online Advertisements Increase Political Candidates' Name Recognition or Favorability? Evidence from Randomized

Field Experiments'. *Political Behavior* 36 (2): 263–89. https://doi.org/10.1007/s11109-013-9239-z.

Cadwalladr, Carole. 2017. 'The Great British Brexit Robbery: How Our Democracy Was Hijacked'. *The Observer*, 7 May 2017, sec. Technology. http://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy.

———. 2018. '"I Made Steve Bannon's Psychological Warfare Tool": Meet the Data War Whistleblower'. *The Guardian*, 18 March 2018, sec. News. http://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump.

Chester, Jeff, and Kathryn C. Montgomery. 2017. 'The Role of Digital Marketing in Political Campaigns'. *Internet Policy Review* 6 (4). https://doi.org/10.14763/2017.4.773.

Cowls, Josh, and Ralph Schroeder. 2018. 'Tweeting All the Way to the White House'. In *Trump and the Media*, edited by Pablo J. Boczkowski and Zizi Papacharissi, 151–58. Cambridge, Massachusetts: MIT Press.

Fowler, Erika Franklin, Michael M. Franz, and Travis N. Ridout. 2015. *Political Advertising in the United States*. Boulder, Colorado: Westview Press.

Furnham, Adrian, and Mark Fenton-O'Creevy. 2018. 'Personality and Political Orientation'. *Personality and Individual Differences* 129 (July): 88–91. https://doi.org/10.1016/j.paid.2018.03.020.

Halpern, Sue. 2017. 'How He Used Facebook to Win'. *The New York Review of Books*, 8 June 2017. http://www.nybooks.com/articles/2017/06/08/how-trump-used-facebook-to-win/.

Harding, James. 2008. *Alpha Dogs: How Political Spin Became a Global Business*. Main edition. New York: Atlantic Books.

Henderson, John A., and Alexander G. Theodoridis. 2017. 'Seeing Spots: Partisanship, Negativity and the Conditional Receipt of Campaign Advertisements'. *Political Behavior*, October, 1–23. https://doi.org/10.1007/s11109-017-9432-6.

Hersh, Eitan. 2015. *Hacking the Electorate: How Campaigns Perceive Voters*. New York, NY: Cambridge University Press.

Hersh, Eitan D., and Brian F. Schaffner. 2013. 'Targeted Campaign Appeals and the Value of Ambiguity'. *The Journal of Politics* 75 (2): 520–34. https://doi.org/10.1017/S0022381613000182.

Howard, Philip N. 2006. *New Media Campaigns and the Managed Citizen*. Cambridge: Cambridge University Press.

Illing, Sean. 2017. 'A Political Scientist Explains How Big Data Is Transforming Politics'. Vox. 16 March 2017. https://www.vox.com/conversations/2017/3/16/14935336/big-data-politics-donald-trump-2016-elections-polarization.

Kalla, Joshua L., and David E. Broockman. 2018a. 'The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments'. *American Political Science Review* 112 (1): 148–66. https://doi.org/10.1017/S0003055417000363.

———. 2018b. 'The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments'. *American Political Science Review* 112 (1): 148–66. https://doi.org/10.1017/S0003055417000363.

Karpf, David. 2017. 'Will the Real Psychometric Targeters Please Stand Up?' Civic Hall. 1 February 2017. https://civichall.org/civicist/will-the-real-psychometric-targeters-please-stand-up/.

Kim, Young Mie, Jordan Hsu, David Neiman, Colin Kou, Levi Bankston, Soo Yun Kim, Richard Heinrich, Robyn Baragwanath, and Garvesh Raskutti. 2018. 'The Stealth Media? Groups and Targets behind Divisive Issue Campaigns on Facebook'. *Political Communication* 0 (0): 1–29. https://doi.org/10.1080/10584609.2018.1476425.

Kreiss, Daniel. 2017. 'Micro-Targeting, the Quantified Persuasion'. *Internet Policy Review* 6 (4). https://doi.org/10.14763/2017.4.774.

Kreiss, Daniel, and Shannon C. McGregor. 2017. 'Technology Firms Shape Political Communication: The Work of Microsoft, Facebook, Twitter, and Google With Campaigns During the 2016 U.S. Presidential Cycle'. *Political Communication* 35 (2): 155–77. https://doi.org/10.1080/10584609.2017.1364814.

Krogerus, Mikael, and Hannes Grassegger. 2017. 'The Data That Turned the World Upside Down'. Motherboard. 28 January 2017. https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win.

Kuchler, Hannah. 2016. 'How "Big Data" Analysts Are Counting on Your Vote'. Financial Times. 27 May 2016. https://www.ft.com/content/fc1879be-1ed4-11e6-b286-cddde55ca122.

Martin, Gregory J., and Ali Yurukoglu. 2017. 'Bias in Cable News: Persuasion and Polarization'. *American Economic Review* 107 (9): 2565–99. https://doi.org/10.1257/aer.20160812.

McGregor, Shannon C., and Natalie Jomini Stroud. 2018. *Digital Discussions: How Big Data Informs Political Communication*. London: Routledge. https://

www.routledge.com/Digital-Discussions-How-Big-Data-Informs-Political-Communication/Stroud-McGregor/p/book/9780815381860.

Meyer, Eric T., and Ralph Schroeder. 2015. *Knowledge Machines: Digital Transformations of the Sciences and Humanities*. MIT Press.

Mitchell, Richard. L. 2012. Political IT: How campaigns mine the Web - part 2. *Computerworld*, 31 October, n.p. Avaliable at: https://www.computerworld.com/article/2492950/big-data/political-it--how-campaigns-mine-the-web---part-2.html

Napoli, Philip M. 2010. *Audience Evolution: New Technologies and the Transformation of Media Audiences*. New York: Columbia University Press.

Neuman, W. Russell. 2016. *The Digital Difference: Media Technology and the Theory of Communication Effects*. Cambridge, Massachusetts: Harvard University Press.

Nickerson, David W., and Todd Rogers. 2014. 'Political Campaigns and Big Data'. *Journal of Economic Perspectives* 28 (2): 51–74. https://doi.org/10.1257/jep.28.2.51.

Nielsen, Rasmus Kleis. 2012. *Ground Wars: Personalized Communication in Political Campaigns*. Princeton, New Jersey: Princeton University Press.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group.

Oremus, Will. 2018. 'The Internet Isn't Why Trump Won, Study Finds'. Slate Magazine. 24 July 2018. https://slate.com/technology/2018/07/the-internet-isnt-why-trump-won-stanford-and-brown-study-finds.html.

Schroeder, Ralph. 2018a. 'Towards a Theory of Digital Media'. *Information, Communication & Society* 21 (3): 323–39. https://doi.org/10.1080/1369118X.2017.1289231.

———. 2018b. *Social Theory After the Internet: Media, Technology, and Globalization*. 1 edition. London: UCL Press.

Simon, Felix M. 2018. 'The Big Data Panic'. *Viewpoints* (blog). 25 March 2018. https://medium.com/viewpoints/cambridge-analytica-and-the-big-data-panic-5029f12e1bcb.

Susskind, Jamie. 2018. *Future Politics*. 1st ed. Oxford: Oxford University Press.

Tufekci, Zeynep. 2014. 'Engineering the Public: Big Data, Surveillance and Computational Politics'. *First Monday* 19 (7). http://firstmonday.org/ojs/index.php/fm/article/view/4901.

Turow, Joseph. 2017. *The Aisles Have Eyes: How Retailers Track Your Shopping, Strip Your Privacy, and Define Your Power*. New Haven London: Yale University Press.

Tworek, Heidi. 2017. 'Cambridge Analytica, Trump, and the New Old Fear of Manipulating the Masses'. *Nieman Lab* (blog). 15 May 2017. http://www.niemanlab.org/2017/05/cambridge-analytica-trump-and-the-new-old-fear-of-manipulating-the-masses/.

Vaidhyanathan, Siva. 2018. *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy*. Oxford, New York: Oxford University Press.

Weaver, Matthew. 2018. 'Social Media "micro-Targeting" of Voters on the Increase, MPs Told'. The Guardian. 23 January 2018. https://www.theguardian.com/media/2018/jan/23/social-media-micro-targeting-of-voters-on-the-increase-mps-told.

Williams, Christine B., and Girish J. "Jeff" Gulati. 2017. 'Digital Advertising Expenditures in the 2016 Presidential Election'. *Social Science Computer Review*, September, 1–16. https://doi.org/10.1177/0894439317726751.

Appendix. The sample of political data analytics companies

Reference Number	Organisation name	Organisation URL
1	Cambridge Analytica (CA Political)	https://ca-political.com/
2	AggregateIQ	https://aggregateiq.com/
3	BlueLabs	http://www.bluelabs.com/
4	Clarity Campaign Labs	http://www.claritycampaigns.com/
5	CivisAnalytics	https://www.civisanalytics.com/
6	TargetedVictory	http://www.targetedvictory.com/
7	Avantgarde Analytics	https://www.avntgrd.com/
8	Echelon Insights	http://echeloninsights.com/
9	Catalist	https://www.catalist.us/
10	Aristotle Inc.	http://aristotle.com/
11	Grassroots Targeting	http://www.grassrootstargeting.com/
12	Revolution Messaging	https://revolutionmessaging.com/
13	dspolitical	https://dspolitical.com/
14	L2 Political	http://www.l2political.com/
15	Resonate	https://www.resonate.com/
16	Yuhas	https://www.yuhasgroup.com/
17	HaystaqDNA	https://haystaqdna.com/
18	WPAi	http://wpaintel.com/
19	Deep Root Analytics	https://www.deeprootanalytics.com/