Data-driven political campaigns in practice: understanding and regulating diverse data-driven campaigns

Katharine Dommett

Department of politics and international relations, University of Sheffield, United Kingdom, k.dommett@sheffield.ac.uk

Abstract: Data-driven campaigning has become a feature of political campaigns around the world. There is growing evidence that political campaigners at the elite and grassroots level believe that data matters for electoral success. This belief is having important consequences for the way that political campaigns are being performed. However, in practice, data-driven campaigning does not take a single form and there are very different practices in the way data is being used. These variations matter because they have different democratic implications. Whilst some uses of data may be viewed as permissible, others can raise democratic concerns. This article casts light on a diversity of different data-driven practices evident in campaigns by offering a theoretical account of the different ways in which data can be used. Discussing three factors that characterise the activities of political campaigners, this article demonstrates variations in who is using data in campaigns, what the sources of campaign data are, and how data informs communication. Reviewing the landscape of data-driven campaigning within political parties, and drawing extensive analysis of the UK case, this article presents categories that can be used to map campaign practice and identify activities to which regulators may wish to respond.

This paper is part of Data-driven elections, a special issue of Internet Policy Review *guest-edited by Colin J. Bennett and David Lyon.*

INTRODUCTION

Data has become an important part of how we understand political campaigns. In reviewing coverage of elections – particularly in the US – the idea that political parties and campaigners now utilise data to deliver highly targeted, strategic and successful campaigns is readily found. In academic and non-academic literature, it has been argued that "[i]n countries around the world political parties have built better databases, integrated online and field data, and created more sophisticated analytic tools to make sense of these traces of the electorate" (Kreiss and Howard, 2010, p. 1; see also in t'Veld, 2017, pp. 2-3). These tools are reported to allow voters to "be monitored and targeting continuously and in depth, utilising methods intricately linked with and drawn from the commercial sector and the vast collection of personal and individual data" (Kerr Morrison, Naik, and Hankey, 2018, p. 11). The Trump campaign in 2016 is accordingly claimed to have "target[ed] 13.5 million persuadable voters in sixteen battleground states, discovering the hidden Trump voters, especially in the Midwest" (Persily, 2017, p. 65). On the basis of such accounts, it appears that data-driven campaigning is coming to define electoral practice – especially in the US - and is now key to understanding modern campaigns.

Yet, at the same time, important questions have been raised about the sophistication and uptake of data-driven campaign tools. As Baldwin-Philippi (2017) has argued, there are certain "myths" about data-driven campaigning. Studying campaigning practices Baldwin-Philippi has shown that "all but the most sophisticated digital and data-driven strategies are imprecise and not nearly as novel as the journalistic feature stories claim" (2017, p. 627). Hersh (2015) has also shown that the data that parties possess about voters is not fine-grained, and tends to be drawn from public records that contain certain standardised information. Moreover, Bennett has highlighted the significant incentive that campaign consultants and managers have to emphasise the sophistication and success of their strategies, suggesting that campaigners may not be offering an accurate account of current practices (2016, p. 261; Kreiss and McGregor, 2018).

These competing accounts raise questions about the nature of data-driven campaigning and the extent to which common practices in data use are found around the globe. These ideas are conceptually important for our understanding of developments in campaigning, but they also have significance for societal responses to the practice of data-driven campaigning. With organisations potentially adopting different data-driven campaigning practices it is important to ask which forms of data use are seen to be democratically acceptable or problematic. [1] These questions are particularly important given the recent interest from international actors and politicians in understanding and responding to the use of data analytics (Information Comissioners Office, 2018a), and specifically practices at Facebook (Kang et al., 2018). Despite growing pressure from these actors to curtail problematic data-driven campaigning practices, it is as yet unclear precisely what is unacceptable and how prevalent these practices are in different organisations and jurisdictions. For these reasons there is a need to understand more about data-driven campaigning.

To generate this insight, in this article I pose the question: "what practices characterise data-driven campaigning?" and develop a comparative analytical framework that can be used to understand, map and consider responses to data-driven campaigning. Identifying three facets of this question, I argue that there can be variations in *who* is using data in campaigns, *what* the sources of data are, and *how* data informs communication in campaigns. Whilst not exhaustive, these questions and the categories they inspire are used to outline the diverse practices that

constitute data-driven campaigning within single and different organisations in different countries. It is argued that our understanding of who, what and how data is being used is critical to debates around the democratic acceptability of data-driven campaigning and provides essential insights required when contemplating a regulatory response.

This analysis and the frameworks it inspires have been developed following extensive analysis of the UK case. Drawing on a three-year project exploring the use of data-driven campaigning within political parties, the analysis discusses often overlooked variations in how data is used. In highlighting these origins I contend that these questions are not unique to the UK case, but can inspire analysis around the globe and in different organisations. Indeed, as I will discuss below, this form of inquiry is to be encouraged as comparative analysis makes it possible to explore how different legal, institutional and cultural contexts affect data-driven campaigning practices. Furthermore, analysis of different kinds of organisation makes it possible to understand the extent to which party practices are unique. Although this article is therefore inspired by a particular context and organisational type, the questions and frameworks it provides can be used to unpack and map the diversity of data-driven campaigning practices, providing conceptual clarity able to inform a possible regulatory response.

DATA AND ELECTION CAMPAIGNS

The relationship between data and election campaigns is well established, particularly in the context of political parties. Describing the focus of party campaigning, Dalton, Farrell and McAllister (2013) outline the longstanding interest parties have in collecting data that can be analysed to (attempt to) achieve electoral success. In their account, "candidates and party workers meet with individual voters, and develop a list of people's voting preferences. Then on election day a party worker knocks on the doors of prospective supporters at their homes to make sure they cast their ballot and often offers a ride to the polls if needed" (p. 56). Whilst parties in different contexts are subject to different regulations and norms that affect the data they can collect and use (Kreiss and Howard, 2010), it is common for them to be provided with information by the state about voters' age, registered status and turnout history (Hersh, 2015). In addition, parties then tend to gather their own data about voter interests, voting preferences and degree of support, allowing them to build large data sets and email lists at national and local levels. Although regulated – most notably through the General Data Protection Regulation (GDPR), which outlines rules in Europe for how data can be collected, used and stored – parties' use of data is often seen to be democratically permissible as it enables participation and promotes an informed citizenry.

In recent history, the use of data by parties is seen to have shifted significantly, making it unclear how campaigns are organised and whether they are engaging in practices that may not be democratically appropriate. In characterising these practices, two very different accounts of data use have emerged. On the one hand, scholars such as Gibson, Römmele and Williamson (2014) have argued that parties now adopt data-driven campaigns that "focus on mining social media platforms to improve their voter profiling efforts" (p. 127). From this perspective, parties are now often seen to be routinely using data to gain information, communicate and evaluate campaign actions.

In terms of *information,* it has been argued that data-driven campaigning draws on new sources of data (often from social media and online sources) to allow parties to search for patterns in citizens' attitudes and behaviours. Aggregating data from many different sources at a level

hitherto impossible, data-driven campaigning techniques are seen to allow parties to use techniques common in the commercial sector to "construct predictive models to make targeting campaign communications more efficient" (Nickerson and Rogers, 2014, p. 54; Castleman, 2016; Hersh, 2015, p. 28). Similarly, attention has been directed to the capacity to use algorithms to identify "look-alike audiences" (Tactical Tech, 2019, pp. 37-69), 2 allowing campaigners to find new supporters who possess the same attributes as those already pledged to a campaign (Kreiss, 2017, p. 5). Data-driven campaigning techniques are therefore seen to offer campaigns additional information with minimal investment of resources (as one data analyst becomes able to find as many target voters as an army of grassroots activists) (Dobber et al., 2017, p. 4).

In addition, data-driven campaigning has facilitated targeted *communication* (Hersh, 2015, pp. 1-2), allowing particular messages to be conveyed to certain kinds of people. These capacities are seen to enable stratified campaign messaging, allowing personalised messages that can be delivered fast through cheap and easy to use online (and offline) interfaces. Data-driven campaigning has therefore been reported to allow campaigners to "allocate their finite resources more efficiently" (Bennett, 2016, p. 265), "revolutioniz[ing] the process" of campaigning (International IDEA, 2018, p. 7; Chester and Montgomery, 2017).

It has also been claimed that data-driven campaigning enables parties to *evaluate* campaign actions and gather feedback in a way previously not possible. Utilising message-testing techniques such as A/B testing, and monitoring response rates and social media metrics, campaigners are seen to be able to use data to analyse – in real time – the impact of campaign actions. Whether monitoring the effect of an email title on the likelihood that it is opened by recipients (Nickerson and Rogers, 2014, p. 57), or testing the wording that makes a supporter most likely to donate funds, data can be gathered and analysed by campaigns seeking to test whether their interventions work (Kreiss and McGregor, 2018, pp. 173-4; Kerr Morrison et al., 2018, p. 12; Tactical Tech, 2019). 3

These new capacities are often highlighted in modern accounts of campaigning and suggest that there has been significant and rapid change in the activities of campaigning organisations. Whilst prevalent, this idea has, however, been challenged by a small group of scholars who have offered a more sceptical account, arguing that "the rhetoric of data-driven campaigning and the realities of on-the-ground practices" are often misaligned (Baldwin-Philippi, 2017, p. 627).

THE SCEPTICAL ACCOUNT

A number of scholars of campaign practice have questioned the idea that elections are characterised by data-driven campaigning and have highlighted a gulf between the rhetoric and reality of practices here. Nielsen, for example, has shown that whilst data-driven tools are available, campaigns continue to rely primarily on "mundane tools" (2010, p. 756) such as email to organise their activities. Hersh also found that, in practice, campaigns do not possess "accurate, detailed information about the preference and behaviours of voters" (2015, p. 11), but rely instead on relatively basic, publically available data points. Similar observations led Baldwin-Philippi to conclude that the day-to-day reality of campaigning is "not nearly as novel as the journalistic feature stories claim" as "campaigns often do not execute analytic-based campaigning tactics as fully or rigorously as possible" (2017, p. 631). In part the gulf between possible and actual practice has emerged because parties – especially at a grassroots level – lack the capacity and expertise to utilise data-driven campaigning techniques (Ibid., p. 631). There is

accordingly little evidence that parties are routinely using data to gain more information about voters, to develop new forms of targeted communication or to evaluate campaign interventions. Indeed, in a study of the UK, Anstead et al. found no evidence "that campaigns were seeking to send highly targeted but contradictory messages to would-be supporters", with their study of Facebook advertisements showing that parties placed adverts that reflected "the national campaigns parties were running" (unpublished, p. 3).

Other scholars have also questioned the scale of data-use by highlighting the US-centric focus of much scholarship on political campaigns (Kruschinski and Haller, 2017; Dobber at al., 2017). Kreiss and Howard (2010) have highlighted important variations in campaign regulation that restrict the practices of data-driven campaigns (see also: Bennett, 2016). In this way, a study of German campaigning practices by Kruschinski and Haller (2017) highlights how regulation of data collection, consent and storage means that "German campaigners cannot build larger data-bases for micro-targeting" (p. 8). Elsewhere Dobber et al. (2017, p. 6) have highlighted how different electoral systems, regulatory systems and democratic cultures can inform the uptake of data-driven campaigning tools. This reveals that, whilst often discussed in universal terms, there are important country and party level variations that reflect different political, social and institutional contexts. [4] These differences are not, however, often highlighted in existing accounts of data-driven campaigning.

Reflecting on reasons for this gulf in rhetoric and practice, some attention has been directed to the incentives certain actors have to "sell" the sophistication and success of data-driven campaigning practices. For Bennett, political and technical consultants "are eager to tout the benefits of micro-targeting and data-driven campaigning, and to sell a range of software applications, for both database and mobile environments" (2016, p. 261). Indeed, with over 250 companies operating worldwide that specialise in the use of individual data in political campaigns (Kerr Morrison, Naik, and Hankey, 2018, p. 20), there is a clear incentive for many actors to "oversell" the gains to be achieved through the use of data-targeting tools (a behaviour Cambridge Analytica has, for example, been accused of). Whatever the causes of these diverging narratives, it is clear that our conceptual understanding of the nature of data-driven campaigning, and our empirical understanding of how extensively different practices are found is underdeveloped. We therefore currently lack clear benchmarks against which to monitor the form and extent of data-driven campaigning.

These deficiencies in our current conceptualisation of data-driven campaigning are particularly important because there has been recent (and growing) attention paid to the need to regulate data-use in campaigns. Indeed, around the globe calls for regulation have been made citing concerns about the implications of data-driven campaigning for privacy, political debate, transparency and social fragmentation (Dobber et al, 2017, p. 2). In the UK context, for example, the Information Commissioner, Elizabeth Denham, launched an inquiry into the use of data analytics for political purposes by proclaiming:

> [w]hat we're looking at here, and what the allegations have been about, is mashing up, scraping, using large amounts of personal data, online data, to micro target or personalise or segment the delivery of the messages without individuals' knowledge. I think the allegation is that fair practices and fair democracy is under threat if large data companies are processing data in ways that are invisible to the public (quoted in Haves, 2018, pp. 2-3).

Similar concerns have been raised by the Canadian Standing Committee on Access to Information, Privacy and Ethics, the US Senate Select Committee on Intelligence, and by international bodies such as the European Commission. These developments are particularly pertinent because the conceptual and empirical ambiguities highlighted above make it unclear which data-driven campaign practices are problematic, and how extensively they are in evidence.

It is against this backdrop that I argue there is a need to unpack the idea of data-driven campaigning by asking "what practices characterise data-driven campaigning?". Posing three supplementary questions, in the remainder of the article I provide a series of conceptual frameworks that can be used to understand and map a diversity of data use practices that are currently obscured by the idea of data-driven campaigning. This intervention aims not only to clarify our conceptual understanding of data-driven campaigning practices, and to provide a template for future empirical research, but also to inform debate about the democratic acceptability of different practices and the form any regulatory response should take.

NAVIGATING THE PRACTICE OF DATA-DRIVEN CAMPAIGNS

Whilst often spoken about in uniform terms, data-driven campaigning practices come in a variety of different forms. To begin to understand the diversity of different practices, it is useful to pose three questions:

1. Who is using data in campaigns?
2. What are the sources of campaign data?
3. How does data inform communication?

For each question, I argue that it is possible to identify a range of answers rather than single responses. Indeed, different actors, sources and communication strategies can be associated with data use within single as well as between different campaigns. Recognising this, I develop three analytical frameworks (one for each question) that can be used to identify, map and contemplate different practices.

These frameworks have been designed to enable comparative analysis between different countries and organisations, highlighting the many different ways in which data is used. Whilst not applied empirically within this article, the ideal type markers outlined below can be operationalised to map different practices. In doing so it should be expected that a spectrum of different positions will be found within any single organisation. Whilst it is not within the scope of this paper to fully operationalise these frameworks, methods of inquiry are discussed to highlight how data may be gathered and used in future analysis. In the discussion below, I therefore offer these frameworks as a conceptual device that can be built upon and extended in the future to generate comparative empirical insights. This form of empirical analysis is vital because it is expected that answers to the three questions will vary depending on the specific geographic or organisational context being examined, highlighting differences in data driven campaigning that need to be recognised by those considering regulation and reform.

WHO IS USING DATA IN CAMPAIGNS?

When imagining the orchestrators of data-driven campaigning the actors that come to mind are often data specialists who provide insights for party strategists about how best to campaign.

Often working for an external company or hired exclusively for their data expertise, these actors have received much coverage in election campaigns. Ranging from the now notorious Cambridge Analytica, to established companies such as BlueStateDigital and eXplain (formerly Liegey Muller Pons), there is often evidence that professional actors facilitate data-driven campaigns. Whilst the idea that parties utilise professional expertise is not new (Dalton et al., 2001, p. 55; Himmelweit et al., 1985, pp. 222-3), data professionals are seen to have gained particular importance because "[n]ew technologies require new technicians" (Farrell et al., 2001). This means that campaigners require external, professional support to utilise new techniques and tools (Kreiss and McGregor, 2018; Nickerson and Rogers, 2014, p. 70). Much commentary therefore gives the impression that data-driven campaigning is being facilitated by an elite group of professional individuals with data expertise. For those concerned about the misuse of data and the need to curtail practices seen to have negative democratic implications, this conception suggests that it is the actions of a very small group that are of concern. And yet, as the literature on campaigns demonstrates, parties are reliant on the activism of local volunteers (Jacobson, 2015), and often lack the funds to pay for costly data expertise (indeed, in many countries spending limits prevent campaigners from paying for such expertise). As a result, much data-driven campaigning is not conducted by expert data professionals.

In thinking through this point, it is useful to note that those conducting data-driven campaigning can have varying professional status and levels of expertise. These differences need to be recognised because they affect both who researchers study when they seek to examine data-driven campaigning, but also whose actions need to be regulated or overseen to uphold democratic norms. [5] Noting this, it is useful to draw two conceptual distinctions between professional and activist data users, and between data novices and experts. These categories interact, allowing four "ideal type" positions to be identified in Figure 1.

Professional	Experts in data-driven campaigning		Volunteer
	Expert data professionals *Cambridge Analytica; Party Data or Targeting Analyst*	**Expert data activists** *Momentum (UK); Sanders campaign (US)[6]*	
	Professionals without data expertise *ECanvasser; Party Campaigns or Membership Officer*	**Activists without data expertise** *Local party volunteer*	
	Novices in data-driven campaigning		

Figure 1: Who is using data in campaigns?[6]

Looking beyond the "expert data professionals" who often spring to mind when discussing data-driven campaigning, Figure 1 demonstrates that there can be different actors using data in campaigns. It is therefore common to find "professionals without data expertise" who are employed by a party. Whilst often utilising or collecting data, these individuals do not possess the knowledge to analyse data or develop complex data-driven interventions. Interestingly, this group has been understudied in the context of campaigns, meaning the precise differences between external and internal professionals are not well understood.

In addition to professionals, Figure 1 also shows that data-driven campaigning is performed by activists who can vary in their degree of expertise. Some, described here as "expert data activists", can possess specialist knowledge - often having many of the same skills as expert data professionals. Others, termed "activists without data expertise", lack even basic understandings of digital technology (let alone data-analysis) (Nielsen, 2012). Some attention has been paid to activists" digital skills in recent elections with, for example, coverage of digital expertise amongst Momentum activists in the UK (Zagoria and Schulkind, 2017) and Bernie Sanders activists in the US (Penney, 2017). And yet, other studies have suggested that such expertise is not common amongst activists (Nielsen, 2012).

These classifications therefore suggest that data-driven campaigning can and is being conducted by very different actors who vary in their relationship with the party, and in their expertise. Currently we have little insight into the extent to which these different actors dominate campaigns, making it difficult to determine who is using data, and hence whose activities (if any) are problematic. This indicates the need for future empirical analysis that sets out to determine the prevalence and relative power of these different actors within different organisations. Whilst space prevents a full elucidation of the markers that could be used for this analysis, it would be possible to map organisational structures and use surveys to gauge the extent of data-expertise present amongst professionals and activists. In turn, these insights could be mapped against practices to determine who was using data in problematic ways. It may, for example, be that whilst "expert data professionals" are engaging in practices that raise questions about the nature of democratic debate (such as micro-targeting), "activists without data expertise" may be using data in ways that raise concerns about data security and privacy.

Knowing who is using data how is critical for thinking about where any response may be required, but also when considering how a response can be made. Far from being subject to the same forms of oversight these different categories of actors are subject to different forms of control. Whilst professionals tend to be subject to codes of conduct that shape data use practices, or can be held accountable by the threat of losing their employment, the activities of volunteers can be harder to regulate. As shown by Nielsen (2012), even when provided with central guidance and protocols, local activists often diverge from central party instructions, reflecting a classic structure/agency dilemma. This suggests not only that the activities of different actors may require monitoring and regulation, but also that different responses may be required. The question "who is using data in campaigns?" therefore spotlights a range of practices and democratic challenges that are often overlooked, but which need to be appreciated in developing our understanding and any regulatory response.

WHAT ARE THE SOURCES OF CAMPAIGN DATA?

Having looked at who is using data in campaigns, it is, second, important to ask what are the sources of campaign data? The presumption inherent in much coverage of data-driven campaigning is that campaigners possess complex databases that hold numerous pieces of data about each and every individual. The International Institute for Democracy and Electoral Assistance (IDEA), for example, has argued that parties "increasingly use big data on voters and aggregate them into datasets" which allow them to "achieve a highly detailed understanding of the behaviour, opinions and feelings of voters, allowing parties to cluster voters in complex groups" (2018, p. 7; p. 5). It therefore often appears that campaigns use large databases of information composed of data from different (and sometimes questionable) sources. However, as suggested above, the data that campaigns possess is often freely disclosed (Hersh, 2015), and many campaigners are currently subject to privacy laws around the kind of data they can collect and utilise (Bennett, 2016; Kruschinski and Haller, 2017).

To understand variations and guide responses, four more categories are identified. These are determined by thinking about variations in the form of data; differentiating between disclosed and inferred data, and the conditions under which data is made available; highlighting differences between data that is made available without charge, and data that is purchased.

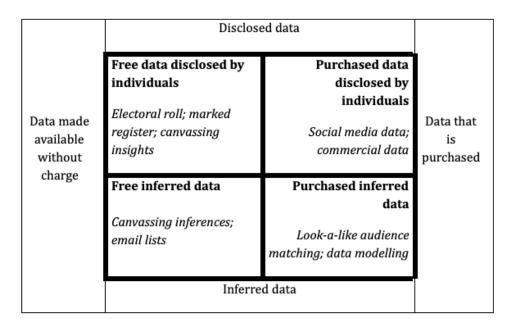|	Disclosed data		
Data made available without charge	**Free data disclosed by individuals** *Electoral roll; marked register; canvassing insights*	**Purchased data disclosed by individuals** *Social media data; commercial data*	Data that is purchased
	Free inferred data *Canvassing inferences; email lists*	**Purchased inferred data** *Look-a-like audience matching; data modelling*	
	Inferred data		

Figure 2: The sources of campaigning data

As described in Figure 2, much of the data that political parties use is provided to them without charge, but it can come in two forms. The first category "free data disclosed by individuals" refers to data divulged to a campaign without charge, either via official state records or directly by an individual to a campaign. The official data provided to campaigns varies from country to country (Dobber et al., 2017, p. 7; Kreiss and Howard, 2010, p. 5) but can include information on who is registered to vote, a voter's date of birth, address and turnout record. In the US it can even include data on the registered partisan preference of a particular voter (Bennett, 2016, p. 265; Hersh, 2015). This information is freely available to official campaigners and citizens are often legally required to divulge it (indeed, in the UK it is compulsory to sign up to the Electoral Register). In addition, free data can also be more directly disclosed by individuals to campaigns through activities such as voter canvassing and surveys that gather data about individuals' preferences and concerns (Aron, 2015, pp. 20-1; Nickerson and Rogers, 2014, p. 57). The second category "free inferred data" identifies data available without charge, but which is inferred rather than divulged. These deductions can occur through contact with a campaign. Indeed, research by the Office of the Information and Privacy Commissioner for British Columbia, Canada describes how party canvassers often collect data about ethnicity, age, gender and the extent of party support by making inferences that the individual themselves is unaware of (2019, p. 22). It is similarly possible for data that campaigns already possess to be used to make inferences. Information gathered from a petition, for example, can be used to make suppositions about an individual's broader interests and support levels. Much of the data campaigners use is therefore available without charge, but differs in form.

In addition, Figure 2 captures the possibility that campaigns purchase data. This data can be classified in two ways. The category "purchased data disclosed by individuals" describes instances in which parties buy data that was not disclosed directly to them, but was provided to other actors. This data can come in the form of social media data (which parties can buy access

to rather than possess), or include data such as magazine subscription lists (Chester and Montgomery, 2017, pp. 3-4; Nickerson and Rogers, 2014, p. 57). Figure 2 also identifies "purchased inferred data". This refers to modelled data whereby inferences are made about individual preferences on the basis of available data. This kind of modelling is frequently accomplished by external companies using polling data or commercially available insights, but it can also be done on social media platforms, with features such as look-a-like audiences on Facebook selling access to inferred data about individuals' views.

Campaigns can therefore use different types of data. Whilst the existing literature has drawn attention to the importance of regulatory context in shaping the data parties in different countries are legally able to use (Kruschinski and Haller, 2017), there are remarkably few comparative studies of data use in different countries. This makes it difficult to determine not only how places vary in their regulatory tolerance of these different forms of data, but also how extensively parties *actually* use them. Such analysis is important because parties' activities are not only shaped by laws, but can also be informed by variables such as resources or available expertise (Hersh, 2015, p. 170). This makes it important to map current practices and explore if and why data is used in different ways by parties around the world. In envisioning such empirical analysis, it is important to note that parties are likely to be sensitive to the disclosure of data sources. However a mix of methods - including interviews with those using data within parties and data subject access requests - can be used to gain insights here.

In the context of debates around data-driven campaigning and democracy, these categories also prompt debate about the acceptability of different practices. Whilst the idea that certain forms of disclosed data should be available without charge is relatively established as an acceptable component of campaigns, it appears there are concerns over the purchase of data and the collection of inferred data. Indeed, in Canada the Office of the Information and Privacy Commissioner for British Columbia recommended that "[a]ll political parties should ensure door-to-door canvassers do not collect the personal information of voters, including but not limited to gender, religion, and ethnicity information unless that voter has consented to its collection" (2019, p. 41). By acknowledging the different sources of data used for data-driven campaigning it is therefore possible to not only clarify what is happening, but also to think about which forms of data can be acceptably used by campaigns.

HOW DOES DATA INFORM COMMUNICATION?

Finally, in thinking about data-driven campaigning much attention has been paid to micro-targeting and the possibility that data-driven campaigning allows parties to conduct personalised campaigns. IDEA has therefore argued that micro-targeting allows parties to "reach voters with customized information that is relevant to them…appealing to different segments of the electorate in different ways" with new degrees of precision (2018, p. 7). In the context of digital politics, micro-targeting is seen to have led parties to:

> …try to find and send messages to their partisan audiences or intra-party supporters, linking the names in their databases to identities online or on social media platforms such as Facebook. Campaigns can also try to find additional partisans and supporters by starting with the online behaviours, lifestyles, or likes or dislikes of known audiences and then seeking out "look-alike audiences", to use industry parlance (Kreiss, 2017, p. 5).

In particular, platforms such as Facebook are seen to provide parties with a "powerful "identity-

based" targeting paradigm" allowing them to access "more than 162 million US users and to target them individually by age, gender, congressional district, and interests" (Chester and Montgomery, 2017, p. 4). These developments have raised important questions about the inclusivity of campaign messaging and the degree to which it is acceptable to focus on specific segments of the population. Indeed, some have highlighted risks relating to mis-targeting (Hersh and Schaffner, 2013) and privacy concerns (Kim et al., 2018, p. 4). However, as detailed above, there are questions about the extent to which campaigns are sending highly targeted messages (Anstead et al., unpublished).

In order to understand different practices, Figure 3 differentiates between audience size; specifying between wide and narrow audiences, and message content; noting differences between generic and specialised messages.

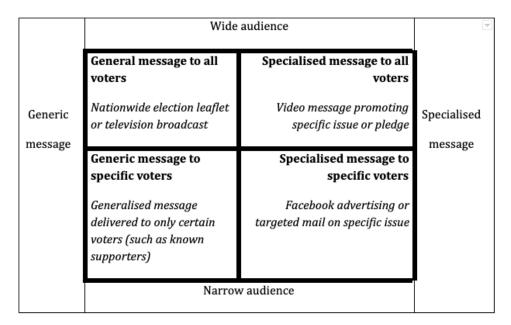|	Wide audience		
Generic message	**General message to all voters** *Nationwide election leaflet or television broadcast*	**Specialised message to all voters** *Video message promoting specific issue or pledge*	Specialised message
	Generic message to specific voters *Generalised message delivered to only certain voters (such as known supporters)*	**Specialised message to specific voters** *Facebook advertising or targeted mail on specific issue*	
	Narrow audience		

Figure 3: How data informs communication

Much campaigning activity comprises generic messages, with content covering a broad range of topics and ideas. By using data (often generated through polling or in focus groups) parties can determine the form of messaging likely to win them appeal. The category "general message to all voters" describes instances in which a general message is broadcast to a wide audience, something that often occurs via party political TV broadcasts or political speeches (Williamson, Miller and Fallon, 2010, p. iii). In contrast "generic message to specific voters" captures instances in which parties limit the audience, but maintain a general message. Such practices often emerge in majoritarian electoral systems where campaigners want to appeal to certain voters who are electorally significant, rather than communicating with (and potentially mobilising) supporters of other campaigns (Dobber et al., 2017, p. 6). Parties therefore often gather data to identify known supporters or sympathisers who are then sent communications that offer a general overview of the party's positions and goals.

Figure 3 also spotlights the potential for parties to offer more specialised messages, describing a campaign's capacity to cover only certain issues or aspects of an issue (focusing, for example, on healthcare rather than all policy realms, or healthcare waiting lists rather than plans to privatise health services). These messages can, once again, be deployed to different audiences. The category "specialised message to all voters" describes instances in which parties use data to

identify a favourable issue (Budge and Farlie, 1983) that is then emphasised in communications with all citizens. In the UK, for example, the Labour Party often communicates its position on the National Health Service, whereas the Conservative Party focuses on the economy (as these are issues which, respectively, the two parties are positively associated with). Finally, "specialised message to specific voters" captures the much discussed potential for data to be used to identify a particular audience that can then be contacted with a specific message. This means that parties can speak to different voters about different issues – an activity that Williamson, Miller and Fallon describe as "segmentation" (2010, p. 6).

These variations suggest that campaigners can use data to inform different communication practices. Whilst much attention has been paid to segmented micro-targeting (categorised here as "specialised messages to specific voters"), there is currently little data on the degree to which each approach characterises different campaigns (either in single countries or different nations). This makes it difficult to determine how extensive different practices are, and whether the messaging conducted under each heading is taking a problematic form. It may, for example, be that specialised messaging to specific voters is entirely innocuous, or it could be that campaigners are offering contradictory messages to different voters and hence potentially misleading people about the positions they will take (Kreiss, 2017, p. 5). Empirically, this form of analysis can be pursued in different ways. As above, interviews with campaign practitioners can be used to explore campaign strategies and targeting, but it is also important to look at the actual practices of campaigns. Resources such as online advertising libraries and leaflet repositories are therefore useful in monitoring the content and focus of campaign communications. Using these methods, a picture of how data informs communication can be developed.

Thinking about the democratic implications of these different practices, it should be noted that message variation by audience size and message scope is not new - campaigns have often varied in their communication practices. And yet digital micro-targeting and voter segmentation has been widely greeted with alarm. This suggests the importance of thinking further about the precise cause of concern here, determining which democratic norms are being violated, and whether this is only occurring in the digital realm. It may, for example, be that concerns do not only reflect digital practices, suggesting that regulation is needed for practices both online and offline. These categories therefore help to facilitate debate about the democratic implications of different practices, raising questions about precisely what it is that is the cause for concern and where a response needs to be made.

DISCUSSION

The above discussion has shown that data-driven campaigning is not a homogenous construct but something conducted by different actors, using different data, adopting different strategies. To date much existing discussion of data-driven campaigning has focused on the extent to which this practice is found. In contrast, in this analysis I have explored the extent to which *different* data-driven campaigning practices can be identified. Highlighting variations in *who* is using data in campaigns, *what* the sources of campaign data are, and *how* data informs campaign communication, I argue that there are a diverse range of possible practices.

What is notable in posing these questions and offering these frameworks is that whilst there is evidence to support these different conceptual categories, at present there is little empirical data on the extent to which each practice exists in different organisations. As such, it is not clear what

proportion of campaign activity is devoted to targeting specific voters with specific messages as opposed to all voters with a general message. Moreover, it is not clear the extent to which parties rely on different actors for data-driven campaigning, nor how much power and scope these actors have within a single campaign. At present, therefore, there is considerable ambiguity about the type of data-driven campaigns that exist. This suggests the urgent need for new empirical analysis that explores the practice of data-driven campaigning in different organisations and different countries. By operationalising the categories proposed here and using methods including interviews, content analysis and data subject access requests, I argue that it is possible to build up a picture of who is using what data how.

Of particular interest is the potential to use these frameworks to generate comparative insights into data-driven campaigning practice. At present studies of data use have tended to be focused on one country, but in order to understand the scope of data-driven campaigning it is necessary to map the presence of different practices. This is vital because, as previous comparative electoral research has revealed, the legal, cultural and institutional norms of different countries can have significant implications on campaigning practices. In this way it would be expected that a country such as Germany with a history of strong data protection law would exhibit very different data-driven campaigning practices to a country such as Australia. In a similar way, it would be expected that different institutional norms would lead a governmental organisation, charity or religious group to use data differently to parties. At present, however, the lack of comparative empirical data makes it difficult to determine what influences the form of data-driven campaigning and how different regulatory interventions affect campaigning practices. This framework therefore enables such comparative analysis, and opens the door to future empirical and theoretical work.

One particularly valuable aspect of this approach is the potential to use these questions and categories to contribute to existing debates around data-driven campaigning and democracy. Throughout the discussion, I have argued that many commentators have voiced concerns. These relate variously to privacy, the inclusivity of political debate, misinformation and disinformation, political finance, external influence and manipulation, transparency and social fragmentation (for more see Zuiderveen Borgesius et al., 2018, p. 92; Chester and Montgomery, 2017, p. 8; Dobber et al., 2017, p. 2; Hersh, 2015, p. 207; Kreiss and Howard, 2010, p. 11; International IDEA, 2018, p. 19). Such concerns have led to calls for regulation, and, as detailed above, many national governments, regulators and international organisations have moved to make a response. And yet, before creating new regulations and laws, it is vital for these actors to possess accurate information about how precisely data-driven campaigning is being conducted, and to reflect on which democratic ideals these practices violate or uphold. Data-driven campaigning is not an inherently problematic activity, indeed, it is an established feature of democratic practice. However, our understanding of the acceptability of this practice will vary dependent on our understanding of who, what and how data is being used (as whilst some practices will be viewed as permissible, others will not). This makes it important to reflect on what is happening and how prevalent these practices are in order to determine the nature and urgency of any regulatory response. Importantly, these insights need to be gathered in the specific regulatory context of interest to policy makers, as it should not be presumed that different countries or institutions will use data in the same way, or indeed have the same standards for acceptable democratic conduct.

The frameworks presented in this article therefore provide an important means by which to consider the nature, prevalence and implications of data-driven campaigning for democracy and can be operationalised to produce vital empirical insights. Such data and conceptual

clarification together can ensure that any reaction to data-driven campaigning takes a consistent, considered approach and reflects the practice (rather than the possibility) of this activity. Given, as a report from Full Fact (2018, p. 31) makes clear that there is a danger of "government overreaction" based on limited information and self-evident assumptions (Ostrom, 2000) about how campaigning is occurring, it is vital that such insights are gathered and utilised in policy debates.

CONCLUSION

This article has explored the phenomenon of data-driven campaigning. Whilst receiving increased attention over recent years, existing debate has tended to focus on the extent to which this practice can be found. In this article, I present an alternative approach, seeking to map the diversity of data-driven campaigning practices to understand the different ways in which data can and is being used. This has shown that far from being characterised by uniform data-driven campaigning practices, data-use can vary in a number of ways.

In classifying variations in *who* is using data in campaigns, *what* the sources of campaign data are, and *how* data informs campaign communication, I have argued that there are diverse practices that can be acceptable to different actors to different degrees. At an immediate level, there is a need to gain greater understanding of what is happening within single campaigns and how practices vary between different political parties around the globe. More widely, there is a need to reflect on the implications of these trends for democracy and the form that any regulatory response may need to take. As democratic norms are inherently contested, there is no single roadmap for how to make a response, but the nature of any response will likely be affected by our understanding of who, what and how data is being utilised. This suggests the need for new conceptual and empirical understanding of data-driven campaigning practices amongst both academics and regulators alike.

REFERENCES

Anstead, N., et al. (2018). *Facebook Advertising the 2017 United Kingdom General Election: The Uses and Limits of User-Generated Data*. Unpublished Manuscript. Retrieved from https://targetingelectoralcampaignsworkshop. files.wordpress.com/2018/02/anstead_et_al_who_targets_me.pdf.

Aron, J. (2015, May 2). Mining for Every Vote. *New Scientist*, *226*(3019), 20–21. https://doi.org/10.1016/S0262-4079(15)30251-7

Baldwin-Philippi, K. (2017). The Myths of Data Driven Campaigning. *Political Communication*, *34*(4), 627–633. https://doi.org/10.1080/10584609.2017.1372999

Bennett, C. (2016). Voter Databases, micro-targeting, and data protection law: can political parties campaign in Europe as they do in North America?. *International Data Privacy Law*, *6*(4), 261–275. https://doi.org/10.1093/idpl/ipw021

Budge, I., & Farlie, D. (1983). *Explaining and Predicting Elections*. London: Allen and Unwin.

Castleman, D. (2016). Essentials of Modelling and Microtargeting. In A. Therriault (Ed.), *Data and Democracy: How Political Data Science is Shaping the 2016 Elections,* (pp. 1–6). Sebastopol, CA: O'Reilly Media. Retrieved from https://www.oreilly.com/ideas/data-and-democracy/page/2/essentials-of-modeling-and-microtargeting

Chester, J., & Montgomery, K.C. (2017). The role of digital marketing in political campaigns. *Internet Policy Review*, *6*(4). https://doi.org/10.14763/2017.4.773

Dalton, R. J., Farrell, D. M., & McAllister, I. (2013). *Political Parties and Democratic Linkage*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199599356.001.0001

Dobber, T., Trilling, D., Helberger, N., & de Vreese, C. H. (2017). Two Crates of Beer and 40 pizzas: The adoption of innovative political behavioral targeting techniques. *Internet Policy Review*, *6*(4). https://doi.org/10.14763/2017.4.777

Dommett, K., & Temple, L. (2018). Digital Campaigning: The Rise of Facebook and Satellite Campaigns. *Parliamentary Affairs*, *71*(1), 189–202. https://doi.org/10.1093/pa/gsx056

Farrell, D., Kolodny, R., & Medvic, S. (2001). Parties and Campaign Professionals in a Digital Age: Political Consultants in the United States and Their Counterparts Overseas. *The International Journal of Press/Politics*, *6*(4), 11–30. https://doi.org/10.1177/108118001129172314

Full Fact. (2018). *Tackling Misinformation in an Open Society* [Report]. London: Full Fact. Retrieved from https://fullfact.org/blog/2018/oct/tackling-misinformation-open-society/

Gibson, R., Römmele, A., & Williamson, A. (2014) Chasing the Digital Wave: International Perspectives on the Growth of Online Campaigning. *Journal of Information Technology & Politics*, *11*(2), 123–129. https://doi.org/10.1080/19331681.2014.903064

Haves, E. (2018). Personal Data, Social Media and Election Campaigns. *House of Lords Library Briefing*. London: The Stationary Office.

Hersh, E. (2015). *Hacking the Electorate: How Campaigns Perceive Voters*. Cambridge: Cambridge University Press.

Hersh, E. & Schaffner, B. (2013). Targeted Campaign Appeals and the Value of Ambiguity. *The Journal of Politics*, *75*(2), 520–534. https://doi.org/10.1017/S0022381613000182

Himmelweit, H., Humphreys, P. , & Jaeger, M. (1985). *How Voters Decide*. Open University Press.

in 't Veld, S. (2017). On Democracy. *Internet Policy Review*, *6*(4). https://doi.org/10.14763/2017.4.779

Information Commissioners Office. (2018a). *Investigation into the use of data analytics in political campaigns*. London: ICO.

Information Commissioners Office. (2018b) *Notice of Intent*. Retrieved from https://ico.org.uk/media/2259363/emmas-diary-noi-redacted.pdf.

International IDEA. (2018). *Digital Microtargeting*. IDEA: Stockholm.

Jacobson, G. (2015). How Do Campaigns Matter?. *Annual Review of Political Science*, *18*(1), 31–47. https://doi.org/10.1146/annurev-polisci-072012-113556

Kaang, C., Rosenburg, M., and Frenkel, S. (2018, July 2). Facebook Faces Broadened Federal Investigations Over Data and Privacy. *New York Times*. Retrieved from https://www.nytimes.com/2018/07/02/technology/facebook-federal-investigations.html?module=inline

Kerr Morrison, J., Naik, R., & Hankey, S. (2018). *Data and Democracy in the Digital Age*. London: The Constitution Society.

Kim, T., Barasz, K., & John, L. (2018). Why Am I Seeing this Ad? The Effect of Ad Transparency on Ad Effectiveness. *Journal of Consumer Research*, *45*(5), 906–932. https://doi.org/10.1093/jcr/ucy039

Kreiss, D., & Howard, P. N. (2010). New challenges to political privacy: Lessons from the first US Presidential race in the Web 2.0 era. *International Journal of Communication,* 4(19), 1032–1050. Retrieved from https://ijoc.org/index.php/ijoc/article/view/870

Kreiss, D. (2017). Micro-targeting, the quantified persuasion. *Internet Policy Review*, *6*(4). https://doi.org/10.14763/2017.4.774

Kreiss, D., & McGregor, S. (2018). Technology Firms Shape Political Communication: The Work of Microsoft, Facebook, Twitter and Google with Campaigns During the 2016 US Presidential Cycle. *Political Communication*, *35*(2), 155–177. https://doi.org/10.1080/10584609.2017.1364814

Kruschinski, S., & Haller, A. (2017). Restrictions on data-driven political micro-targeting in Germany. *Internet Policy Review*, *6*(4). https://doi.org/10.14763/2017.4.780

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

Nielsen, R. (2010). Mundane internet tools, mobilizing practices, and the coproduction of citizenship in political campaigns. *New Media and Society*, *13*(5), 755–771. https://doi.org/10.1177/1461444810380863

Nielsen, R. (2012). *Ground Wars*. Princeton: Princeton University Press.

Office of the Information and Privacy Commissioner for British Columbia. (2019). *Investigation Report P19-01, Full Disclosure: Political Parties, Campaign Data and Voter Consent*. Retrieved from https://www.oipc.bc.ca/investigation-reports/2278

Ostrom, E. (2000). The Danger of Self-Evident Truths. *Political Science and Politics*, *33*(1), 33–44. https://doi.org/10.2307/420774

Penney, J. (2017). Social Media and Citizen Participation in "Official" and "Unofficial" Electoral Promotion: A Structural Analysis of the 2016 Bernie Sanders Digital Campaign. *Journal of Communication*, *67*(3), 402–423. https://doi.org/10.1111/jcom.12300

Persily, N. (2017). Can Democracy Survive the Internet?. *Journal of Democracy*, *28*(2), 63–76. https://doi.org/10.1353/jod.2017.0025

Tactical Tech. (2019). *Personal Data: Political Persuasion – Inside the Influence Industry. How it works*. Berlin: Tactical Technology Collective.

Williamson, A., Miller, L., & Fallon, F. (2010). *Behind the Digital Campaign: An Exploration of the Use, Impact and Regulation of Digital Campaigning*. London: Hansard Society.

Zagoria, T. and Schulkind, R. (2017). How Labour Activists are already building a digital strategy to win the next election", *New Statesman*. Retrieved from https://www.newstatesman.com/politics/elections/2017/07/how-labour-activists-are-already-building-digital-strategy-win-next.

Zuiderveen Borgesius, F., Möller, J., Kruikemeier, S., Fathaigh, R., Irion, K., Dobber, T., Bodo, B. & de Vreese, C. H. (2018). Online Political Microtargeting: Promises and Threats for Democracy. *Utrect Law Review*, *14*(1), 82–96. https://doi.org/10.18352/ulr.420

FOOTNOTES

1. This question is important because it is to be expected that universal responses to this question do not exist, and that different actors in different countries will view and judge practices in different ways (against different democratic standards).

2. See the report from Tactical Tech (2019) *Personal Data* for a range of examples of how data can be used to gain "political intelligence" about voters.

3. Importantly, this data use is not guaranteed to *persuade* voters. Campaigns can identify the type of campaign material viewers are more likely to watch or engage with, but this does not necessarily mean that those same viewers are persuaded by that content.

4. Similarly there are likely to be variations between parties and other types of organisation such as campaign groups or state institutions.

5. It should be noted that these democratic norms are not universal, but are expected to vary dependent on context and the perspective of the particular actor concerned.

6. For more on local expert activism in the UK see Dommett and Temple, 2017. In the US see Penney, 2017.