Towards a critical data science – the complicated relationship between data and the democratic project.

What is driving the rise in data-driven techniques used by politicians and political campaigns to connect with the concerns and needs of citizens? Will a data-driven approach to political campaign messaging disrupt the "echo chamber" effect that is perceived to emerge within online spaces? **Jo Bates** *finds the role of data science in the development of the democratic process is still far from certain.*

Politicians and policy makers are increasingly turning to data science techniques to capture, analyse and respond to the public mood. From the data related controversies of the US primary elections, to the adoption of new data-driven campaigning techniques in the UK's 2015 general election and the Cabinet Office's experiments with data driven Open Policy Making, to the Chinese government's Social Credit System – data science techniques are being developed and used by political institutions in an effort to understand citizens better and, in some cases, nudge peoples' behaviour towards desired ends – from voting decisions to building a "culture of sincerity".

There is some evidence to suggest that as a communication platform the Internet can be used as a tool of persuasion by targeting political content at specific audiences – a targeting that is becoming increasingly fine-grained as a result of new data science techniques. Some data scientists and other observers are hopeful that the techniques being developed can contribute to the improvement of the democratic process by helping politicians to understand voters better and tailor more persuasive personalised campaign messages that reflect individuals' identities, lifestyles and views on a range of issues from immigration to 'gay marriage'.

Image credit: An ALF-CIO canvasser in the 2008 American election – edited CC BY Wikimedia

However, others are concerned that such techniques may contribute to the deepening surveillance of citizens, the manipulation of public discourse by political elites, the embedding of potentially discriminatory forms of social profiling into political communication, and the perception that citizens' role in the democratic process is akin to that of a consumer – factors which ultimately could restrict opportunities for open and informed democratic debate and practice. There is also the significant question of the legality of some data-driven political campaigning methods that are being imported from the USA where personal data protection legislation is much weaker than in the European Union.

Despite these concerns, optimists argue that a data-driven personalised approach to political campaign messaging can disrupt the "echo chamber" effect that is perceived to emerge within online spaces. However, the role of data science in the development of the democratic process is still far from certain. Whilst exposure to a wide range of ideas and debates is crucial to the functioning of a healthy democracy, it is important that our analysis of the democratic potential of these new techniques begins by asking – who is using them to persuade who about what?

"Echo chambers" are commonplace within politics. They are not unique to the political cultures of citizens' Facebook feeds – they go to the heart of our political system. Many would agree with the observation that large sections of our political and media establishment exist in something of an echo chamber of their own – one whose views of the world tend to favour the interests of a privileged few. Whilst debate and disagreements exist within these "echo chambers" they tend to be restricted to the confines of a fairly narrowly defined set of assumptions about how the world does and should work. The response to any fracturing of this "echo chamber" – for example, Corbyn's recent victory in the Labour leadership contest – highlights the depth of desire amongst the establishment for political culture to be constrained to a limited set of ideas about means and ends.

Whilst some politicians may genuinely be driven to use new data-driven techniques to connect with the concerns and needs of citizens, it is clear that generally speaking politics is more complicated than that. In liberal states such as the UK, political elites work hard to persuade citizens to accept the general assumptions of their "echo chamber" and to convince voters that – often despite significant evidence to the contrary – their particular view of how things should be managed is better than their opponents. It is with this need to persuade in mind that politicians are turning to new data science techniques such as personalised campaign messaging.

The fine-grained targeting of messages aimed at convincing the recipient to choose your product is a marketing technique. The adoption of such methods, rather than embracing and enhancing the democratic process, contributes to the remaking of the citizen as a consumer of a political product and empowers political elites in their efforts to draw citizens into accepting the assumptions of their "echo chamber". Imagining and treating citizens as consumers restricts our democratic imagination – it constrains what active citizenship can and should be about.

Image credit: Ilya – Silence CC BY-SA

As well as contributing to techniques of persuasion, the data science gaze also risks emphasising and reproducing majoritarian and populist positions, and erasing more marginalised voices. In order to understand the complexity of political discourse and sentiment we need to seek out and build meaningful, ethical relationships with less visible voices – not lose them to the crowd. In a democratic state, marginalised voices often need to be carefully amplified as their words are not immediately heard and understood by those whose experience of the world is more privileged.

Whilst some data science techniques can contribute to the amplification of marginalised voices it is important that this is done with the desire, consent and, ideally, participation of individuals being drawn attention to – this should be a fundamental principle of a free and democratic society. It would be naïve to assume a benign relationship between political (and other) institutions and ordinary citizens. Powerful institutions can and do oppress and discriminate on the basis of political ideology, sexual life, religious belief, and so on – to forget this erases the struggles of those who are subject to their gaze. It is therefore crucial that as data science techniques become more embedded in a variety of institutional contexts, we protect the right of ordinary citizens to navigate freely the territory between openness and secrecy – because that is the space where marginalised people begin to feel empowered and it is where progressive forms of resistance begin to emerge.

New data science techniques offer immense potential for scientific advancement and human development – there may even be a role for data science in advancing the democratic project. However, in order to ensure that these advances benefit all, rather than empower the few, it is crucial that data scientists work collaboratively with others to incorporate an analysis of power into their practice. We are beginning to see the emergence of exciting work in this area – for example, Catherine D'Ignazio's recent essay imagining what a feminist ethics and politics of data visualization might look like, and the development of new fora such as the Responsible Data Forum that bring together practitioners, activists and academics to tackle such issues.

Data Scientists are also beginning to collaborate with social scientists and humanities scholars to develop innovative teaching programmes such as the MSc Data Science at the Sheffield Information School. It is through such interdisciplinary and cross-sector collaboration that together we can begin to explore the question of the democratic potential of new data science techniques, and work to develop forms of data practice that incorporate questions of power, ethics and agency at their core.

About the Author

Dr Jo Bates *is a Lecturer in Information Politics and Policy based in the University of Sheffield's Information School. Her research explores the socio-cultural factors shaping developments in the political economy of data, with a particular focus on the production and distribution of public datasets and their re-use by third parties including citizens and businesses.*