Big Data and Democracy

Edited by
KEVIN MACNISH AND
JAI GALLIOTT

EDINBURGH
University Press
CONTENTS

Contributors / vii

An Introduction to Big Data and Democracy / 1
Kevin Macnish and Jai Galliott

PART ONE
ONE / Big Data, Consequentialism and Privacy / 13
Kieron O’Hara
TWO / Politics, Big Data and Opacity Respect / 27
Carl Fox
THREE / A Pre-Occupation with Possession: the (Non-)Ownership of Personal Data / 42
Kevin Macnish and Stephanie Gauttier
FOUR / Policing with Big Data: DNA Matching vs Crime Prediction / 57
Tom Sorell

PART TWO
FIVE / Dark Advertising and the Democratic Process / 73
Joe Saunders
SIX / Twitter and Electoral Bias / 89
Wulf Loh, Anne Suphan and Christopher Zirnig
SEVEN / Gated Communities of the Digitised Mind / 104
Thorsten Bronholt
EIGHT / The Network and the Demos: Big Data and the Epistemic Justifications of Democracy / 119
Dave Kinkead and David M. Douglas
PART THREE

NINE / The Technics of a Gnostic World: an Ontogeny of Big Data / 137
John MacWillie

TEN / Trust and Algorithmic Opacity / 153
Steve McKinlay

ELEVEN / Opacity, Big Data, Artificial Intelligence and Machine Learning in Democratic Processes / 167
Ramón Alvarado

TWELVE / The Big Data Paradox and Its Importance to Strategy and Military Systems Development / 182
Tim McFarland, Jai Galliott and Massimiliano Cappuccio

PART FOUR

THIRTEEN / Beyond the Concept of Anonymity: What is Really at Stake? / 201
Björn Lundgren

FOURTEEN / Big Data Analytics and the Accessibility of Public Inquiries / 217
Philip Garnett and Sarah M. Hughes

FIFTEEN / Developing an Ethical Compass for Big Data / 231
Harald Stelzer and Hristina Veljanova

Index / 247
CONTRIBUTORS

Ramón Alvarado is an Assistant Professor of Philosophy and Data Ethics at the University of Oregon. His research focuses on the epistemic and ethical implications of the use of computers and computational methods in science and society.

Thorsten Brønholt conducts research in the Department of Political Science, University of Copenhagen, a school known for its heavy focus on methodologies. He is also a trained project manager. As both an internal and an external consultant, he has designed and carried out several large ($n > 6,000$) quantitative surveys for the Municipality of Copenhagen.

Massimiliano (Max) Cappuccio is Deputy Director of the Values in Defence and Security Technology group. He also holds a secondary affiliation with UAE University. His research in embodied cognitive science and social cognition theory attempts to combine enactive and extended approaches to the mind. As a cognitive philosopher, his research on cognitive processes and intelligent systems is interdisciplinary, and aims to integrate phenomenological analyses, empirical experimentation and synthetic modelling.

David Douglas researches with the Department of Philosophy and is ethics adviser within the Centre for Telematics and Information Technology (CTIT) at the University of Twente, the Netherlands.

Carl Fox is Lecturer in the IDEA Centre at the University of Leeds. He holds a strong interest in the ethics of the media, in particular the relationship between legitimacy and stability in the public sphere. He is currently a co-investigator on an AHRC-funded project examining media regulation and the role of codes of ethics for the press in ten European countries.
Jai Galliott leads the Values in Defence and Security Technology Group within the University of New South Wales at the Australian Defence Force Academy. He is a former Royal Australian Navy officer and Australian Army Research Fellow, and Visiting Fellow at the Centre for Technology and Global Affairs at the University of Oxford.

Philip Garnett is Lecturer in the York Cross-Disciplinary Centre for Systems Analysis & School of Management at the University of York. He is interested in systems theory, complex systems theory, and network analysis techniques that allow us to understand organisations and how they operate. His research is broadly focused on modelling organisational structure and interactions between organisations and their environment, and understanding the flow of information through organisations.

Stephanie Gauttier is a Marie Curie Postdoctoral Fellow at the University of Twente. She has published widely on the use of technological enhancements and wearable technologies in medicine and the workplace. Her current research, sponsored by the Marie Curie IF fellowship looks at how the concept of human enhancement can be translated to organisations, partly by looking at how human enhancement technologies can impact workflow management and the ethical issues this raises.

Sarah M. Hughes is an ESRC-funded Research Fellow in the Department of Geography at the University of Durham. Her current research focuses upon conceptualisations of resistance within the UK asylum system.

David Kinkead teaches in the Department of Philosophy at the University of Queensland.

Wulf Loh researches in the Department of Philosophy at the University of Stuttgart.

Björn Lundgren is researcher in the Division of Philosophy at the Royal Institute of Technology (KTH), Stockholm, Sweden.

Kevin Macnish is Assistant Professor in Ethics and IT at the University of Twente. His research is in the ethics of surveillance, security and technology, and he has published a book and a number of articles on surveillance ethics.

John MacWillie researches at California State University, East Bay, California (CSUEB) with expertise in artificial intelligence, history of science and philosophy of science.

Tim McFarland is a Research Fellow in the Values in Defence and Security Technology group within the School of Engineering and Information Technology of the University of New South Wales at the Australian Defence Force Academy.
Force Academy. His current work is on the social, legal and ethical questions arising from the emergence of new military and security technologies, and their implications for the design and use of military systems.

**Steve McKinlay** is Senior Lecturer in Computing and Research Coordinator at the Wellington Institute of Technology.

**Kieron O’Hara** is Associate Professor in Electronics and Computer Science at the University of Southampton, partly supported by the EPSRC project SOCIAM.

**Joe Saunders** is an Assistant Professor in Philosophy at Durham University, UK. He is the co-editor, with Carl Fox, of *Media Ethics, Free Speech, and the Requirements of Democracy* (2019).

**Tom Sorell** is Professor of Politics and Philosophy and Head of the Interdisciplinary Ethics Research Group in PAIS. He was an RCUK Global Uncertainties Leadership Fellow (2013–16). Previously, he was John Ferguson Professor of Global Ethics and Director of the Centre for the Study of Global Ethics, University of Birmingham.

**Harald Stelzer** is Professor of Political Philosophy and Head of the Section Political Philosophy at the University of Graz, Austria. His main research interests are in the problems and opportunities of the normative evaluation of political courses of action under special consideration of uncertainty and risks in regard to emerging technologies.

**Anne Suphan** is a post-doctoral researcher at Hohenheim University, Germany. She received her PhD in Organization Studies and Cultural Theory from the University of St. Gallen, Switzerland. Her research interests include social media, digital divide, digital sociology and visual research methods.

**Hristina Veljanova** is researcher at the Section Political Philosophy at the University of Graz. She is also pursuing her PhD on designing for values in autonomous vehicles at the University of Graz. Her research interests lie in the areas of philosophy and ethics of technology, machine ethics, robot ethics, moral philosophy and justice theories.

**Christopher Zirnig** is a researcher in the Institute of Law and Social Sciences at the University of Hohenheim.
An Introduction to Big Data and Democracy

Kevin Macnish and Jai Galliott

Over the past decade, political systems across Europe, Asia and North America have been challenged by events including large waves of immigration, a resurgence of economic protectionism and the rise of a pernicious brand of nationalism. Established politicians have adapted their rhetoric and processes in ways that risk alienating the population bases they serve, while populist politicians promising easy solutions have gained significant traction with the electorate. Yet in the coming decade, the greatest challenge to face democracies will be to develop new capabilities and protections that allow for the appropriate management and exploitation of big data. We use the term ‘big data’ to mean the exponentially increasing amount of digital information being created by new information technologies (such as mobile Internet, cloud storage, social networking and the ‘Internet of Things’ that has come to include everything from polling machines to sex toys), as well as the advanced analytics used to process that data. Big data yields not simply a quantitative increase in information, but a qualitative change in how we create new knowledge and understand the world, bringing with it a host of ethical questions. These data-related information technologies have already begun to revolutionise commerce and science, transforming the economy and acting as enablers for other game-changing technology trends, from next-generation genomics to energy exploration.

Despite this, the impact of big data on democracy has only recently come under the microscope. This is not altogether surprising. The term ‘big data’ entered general public consciousness early in 2012 when an article by Charles Duhigg in the New York Times reported on a US superstore identifying a teenage girl as pregnant before she had told her father (Duhigg 2012). The early public focus on the impact of big data thus fell on retail opportunities and
threats to civic and individual values such as privacy. In the UK at least, the Care.Data fiasco of 2014, in which the government announced that all data collected on patients in the UK would be entered into a single database without seeking consent, turned attention to medical applications (Knapton 2016). This was reinforced, again in the UK, with the Google DeepMind/Royal Free Trust hospitals case the same year, in which a private company was being passed sensitive medical information by a public institution (Powles and Hodson 2017; Hern 2017; Suleyman and King 2017).

In September 2016, the CEO of Cambridge Analytica, Alexander Nix, gave a presentation at the annual Concordia Summit in New York City. His presentation, ‘The Power of Big Data and Psychographics’, described his approach to audience targeting and data modelling, and explained how, when combined with psychological profiling, this would ‘enhance’ elections and disrupt conventional marketing (Concordia 2016). Proceeding to discuss how his company had assisted the unsuccessful bid of US Senator Ted Cruz for the Republican Party nomination for the 2016 US Presidential election, Nix explained that his company had developed a model to ‘predict the personality of every single adult in the United States of America’ based on ‘hundreds and hundreds of thousands’ of responses by Americans to a survey. The purpose of this model was to tailor political messages to target individuals with certain personality traits. As Nix explained, ‘today communication is becoming ever increasingly targeted. It’s being individualised for every single person in this room . . . [Y]ou’ll only receive adverts that not only are on the products and services, or in the case of elections, issues that you care about most, but that have been nuanced in order to reflect the way you see the world’ (Concordia 2016).

While Cambridge Analytica would later close after intense scrutiny as to how they had acquired data via a personality test app that an independent researcher had deployed on Facebook (Solon and Laughland 2018), the techniques Nix described demonstrate the possibilities with regard to political messaging on offer thanks to the combination of social media and big data analytics. Facebook in particular faces what some are calling an ‘existential crisis’ over revelations that its user data fell into the hands of the Trump campaign (Byers 2018). Whether the attacks on the social media giant are justified, the fact is that the Obama campaign also employed big data analytic techniques on Facebook data in 2012. Indeed, the re-election campaign for President Barack Obama in the 2012 US Presidential election was seen as innovative precisely for its use of social media and big data analytics (O’Neil 2016: 188–93). However, the reaction from the pundits and press in 2012 was somewhat different, with articles titled ‘Obama, Facebook and the Power of Friendship: the 2012 Data Election’ (Pilkington 2016).
and Michel 2012) and ‘How Obama’s Team Used Big Data to Rally Voters’ (Issenberg 2012). The Obama campaign encouraged supporters to download an Obama 2012 Facebook app that, when activated, let the campaign collect Facebook data both on users and their friends. The campaign later boasted that more than a million people had downloaded the app, which, given an average friend-list size of 190, means that as many as 190 million had at least some of their Facebook data vacuumed up by the Obama campaign without their explicit knowledge or consent (Leetaru 2018). While the use of data analytics was prevalent by both the Obama campaign and the Trump campaign, a significant factor in the meantime has been the rise of people accessing news articles via social media, and the associated rise of ‘fake news’ stories and websites.

In the same year as Trump won the US presidential elections, in the UK the Leave campaign in the European Referendum similarly boasted success, with the official Leave campaign, Vote Leave, proclaiming that big data had played the pivotal role (Cummings 2016). Even more remarkable, is that few people realised that this technology, the collection, aggregation and application of big data, was being utilised during the referendum. It was later revealed that Cambridge Analytica provided initial help and guidance to the Leave.EU campaign, which then went on to develop its own artificial intelligence analysis methodology (Ram 2018). The 2017 elections in Germany were subsequently shrouded in anxiety, some stemming from concerns about the role of big data. Fake accounts and social bots, filter bubbles and echo chambers, foreign propaganda machineries and campaign micro-targeting called the neutrality, inclusiveness and permeability of Germany’s digital public spheres into question (Essif 2017; Kruschinski and Haller 2017).

Big data has meant that political organisations now have the tools to identify issues important to specific individuals, determine how to present their case in the most effective way for those individuals, and the means to communicate their message directly to them. In many ways, this is ideal for the democratic process: if I as a citizen do not care about military spending, why waste my time with advertising about that? Tell me what your party will do to social welfare spending. Big data allows you to know that I care for social welfare and not the military, and allows your party to reach me directly. At the same time, your party can inform a soldier that her job is secure if you should gain power.

The potential downside of this is the result of the soldier and I seeing different messages. Political discourse thus risks becoming a private (typically one-way) conversation between the speaker and the targeted audience. Yet this has always been true: from voting in ancient Athens to politicians
ay, those seeking election could fine-tune their message to individual voters. Likewise, political advertising is hardly new, with some advertisements appealing to some voters and other advertising appealing to different voters. Different posters may go up in different parts of the same city or country depending on the demographic being reached. Is there really anything new in what has been happening with big data in democratic processes?

The chapters in this volume propose several areas of concern which have arisen in recent years in relation to big data and democracy, and which can be conceived as fundamentally novel. These range from trust in automated systems operating on the basis of big data to the aforementioned individualisation of political messaging. In the case of the latter, the substantive change in the nature of advertising is that whereas a variety of political posters may have been used, as posters these were in the public space and available for examination by all citizens, even if they were more likely to be seen by only a sub-set. Individuated political advertising on social media micro-targets the message to a sub-set without the opportunity of others seeing what is being sent.

Any attempt at making political discourse private risks undermining the public discussion of political views and identities. These risks are further exacerbated by the proliferation of echo chambers and filter bubbles, as social media contributes to reinforcing our own prejudices while shielding us from the values, thoughts and feelings of others. As discussed in several chapters of this volume, a stable democracy requires a shared identity and political culture to prevent it fracturing into separate political communities. This in turn necessitates some common communication network for political messages to be transmitted, understood and evaluated by citizens. Without neutral, inclusive and open public spheres, the aggregation of individual interests or beliefs about the common good risks distortion, the control of political institutions loses effectiveness, and the civic commitments of citizens may deteriorate as trust declines.

We are currently in the middle of a technological disruption that will transform the way society is organised and may go on to impact the freedoms of citizens if we do not immediately investigate the challenges posed by big data. The effectiveness of these technologies and techniques appears to have reached new heights, with the digital upheaval in full swing and showing no signs of slowing. The quantity of data we produce doubles every year, so that in 2016 we produced as much processable data as in the history of humankind up to 2015. Every minute we produce hundreds of thousands of Google searches and Facebook posts. Each of these contains information that reveal how we think and feel. The things around us, from
clothing to cars to kettles, are increasingly connected to the Internet. It is estimated that in 10 years’ time there will be 150 billion networked measuring sensors, twenty times more than the number of people on Earth. Then, the amount of data will double every 12 hours.

At the same time, the challenges to democratic society presented by data analytics are not limited to the party-political process. The Snowden revelations included claims that vast quantities of internet data were being collected by the US, UK and other intelligence agencies with the purported goal of uncovering patterns of behaviour which might indicate military or security threats. Artificial intelligence based on assessments of large data sets is increasingly being tested for parole boards and in predictive policing, leading to concerns that people may soon be apprehended for ‘pre-crime’. Furthermore, while many in society have a degree of control over at least some of the data they share, others who rely on welfare do not and must share openly with the state to receive support (Macnish 2017). This volume investigates the phenomenon that is big data and provides a much-needed critical exploration of the extent to which the exponentially increasing amount of digital information available to political actors is transforming the shaping of societies through the democratic process and questions whether there reasonable democratic limits that can be placed on the employment of big data depending on the collection, storage, processing or use of that data.

The book is divided into four parts. Part One looks at the relationship that exists between citizens and data. In the first chapter, Kieron O’Hara applies a Weberian analysis to developments in big data. He argues that just as Weber noted a move from the pre-modern to the modern with the advent of bureaucracy, so we are now entering a time of digital modernity. The focus in digital modernity is less the present (what is happening now) as the subjunctive (what could happen), governed by data. This, he argues, gives data a central role in governing for both good and ill.

In his chapter, Carl Fox argues that citizens deserve ‘opacity respect’ in which the state refrains from peering into citizens’ lives to determine their capacities, for good or ill. Governments should assume that we are rational agents, barring certain obvious exceptions such as infants and people who are mentally incapacitated, in the same way that we assume this of each other. To dig deeper, without permission, is to undermine human dignity and that which renders us equal as humans, thereby forming a foundation of the democratic ideology.

Kevin Macnish and Stephanie Gauttier take a different angle on the question of democratic implications of data by looking at the concept of ownership of personal data. This has been suggested as a way of
giving citizens control of ‘their’ data and is assumed in European law. However, Macnish and Gauttier argue that there are, at best, only weak philosophical grounds on which to base this claim, and that the consequences of viewing our relationship with the data that describe us in terms of ownership may turn out to be highly undesirable. Instead, they argue that control should be based around custody of data and the potential for harm.

Tom Sorell turns our attention to two state (police) uses of big data that have elicited concern: the creation of DNA databases and the use of past data to predict future crimes and criminals. In response to the former, Sorell argues that there is nothing intrinsically wrong with large-scale, indiscriminate databases of DNA profiles. These do not constitute an invasion of privacy, and nor do they necessarily render an entire population suspect, although he accepts that in the current climate they may be interpreted that way. As regards predictive policing, Sorell’s argument is that these uses are more concerning, basing future decisions on past information that may no longer be pertinent and could well be discriminatory.

Part Two turns to look at political advertising, a point of reference in Fox’s chapter, and the centre of attention for Joe Saunders. What is it about the micro-targeting of political advertising, asks Saunders, that is so wrong? His response is that, like dog-whistle politics, sending different advertising to different targets can obscure the open public discussion of policy that is critical to the democratic process. As such, parties (and individuals) that win power through such means find themselves with no democratic mandate to govern.

This discussion is followed by a chapter on Twitter and electoral bias by Wulf Loh, Anne Suphan and Christopher Zirnig. Here the authors introduce Habermasian criteria for the functioning of a healthy democracy and apply research from the use of Twitter during the most recent German general election to describe the influence (or lack thereof) of social media on the electoral process. In particular, the authors highlight what they call a social media divide.

In his chapter, Thorsten Brønholt introduces the concept of gated communities of the digitised mind. These draw on the notions of echo chambers and filter bubbles to suggest that there are regions for at least some of us that function in the same way as a gated community in which we mix only with those least likely to challenge our views. He supports this argument with original analysis of fifteen Danish politicians which summarises the results from semi-structured interviews with the respondents, and an analysis of their personal Facebook and Twitter feeds, as well as identical Google searches on their private devices.
In the final chapter of this section, Dave Kinkead and David Douglas draw on the history of democracies to see how big data and its use with social media sites introduces new challenges to the contemporary marketplace of ideas. They note that traditionally one could narrowcast a tailored message with some impunity, but limited effect, while broadcasts (with larger impact) were open to examination by the public. Micro-targeted political advertising now allows for the narrowcast message to be tweaked and directed on a scale never before seen.

Part Three looks at more technical issues relating to big data. John MacWillie’s chapter develops an ontological understanding of the infrastructure underlying big data applications through an historical overview of developments in information communications technology since the 1950s. This leads him to conclude that big data is a fundamentally new object in the world, bringing with it key issues of richness and complexity in computer networks.

In his chapter, Steve McKinlay argues that the use of big data algorithms introduces a key problem in terms of epistemic opacity. Opacity in various forms is an issue that many authors identify as posing problems for democratic functioning and accountability. In McKinlay’s case, the argument focuses on the impact that epistemic opacity has on our ability to trust non-human agents. He holds that while the outputs of big data-derived decisions can be significant for citizens, where we do not have the ability to understand how these decisions were made we cannot ultimately trust the decider. Decisions based on mere probability are not, he argues, sufficiently grounded for democratic systems and risk harming citizens.

Ramón Alvarado also looks at issues of opacity, but whereas McKinlay argues that the core challenges to arise from this are those relating to trust, Alvarado turns his attention elsewhere. Following a detailed examination of different kinds of opacity, he raises the problem of ‘many hands’ (where to attribute responsibility in complex systems), error assessment and path complexity. In the process, he successfully offers the reader a demystified understanding of how big data computational methods function and suggests ways in which opacity threatens fundamental elements of the democratic process.

Finally in this section, Tim McFarland, Jai Galliott and Massimiliano Cappuccio consider the use of big data in military contexts. They draw on three paradoxes raised by Neil Richards and Jonathan King to examine particular challenges facing the military. These are the transparency paradox, the need for the collection and use of data to be as transparent as possible while being collected and used for national security or military purposes which themselves require secrecy; the identity paradox, which recognises
that while big data reflects the identity and behaviour of those whose data is used, so too can it be used to alter those identities and behaviours; and the power paradox, highlighting the increased power gained through big data coupled with the risks of using those data in a largely unregulated environment.

The fourth and final section of the book takes up the implications for trust in society raised by McKinlay (above). The first chapter in this section, by Björn Lundgren, looks at the meaning and role of anonymity, which is often threatened when data sets are agglomerated in big data practices, leading to de-anonymisation. He argues that there are a number of key values threatened by the de-anonymisation, but that the concept of anonymity is not sufficient given what is really at stake. Instead, he holds that what is really under threat is our ability to be anonymous, which he characterises as a reasonable control over what we communicate.

Philip Garnett and Sarah Hughes then turn their attention to the role of big data in accessing information from public inquiries. Looking in particular at the Chelsea Manning court martial in the US and the Leveson Inquiry in the UK, they argue that the manner in which information pertaining to inquiries is made public is, at best, unsatisfactory. They propose a variety of means to make this information more accessible and hence more transparent to the public through employing big data techniques.

Finally, Harald Stelzer and Hristina Veljanova argue in their chapter for a new ethical compass with which to approach big data concerns. They identify key ethical concerns which often arise in cases regarding big data and then provide a framework through which we might approach these concerns such that we can have a degree of certainty that we have not overlooked ethical worries.

Our hope is that this volume will present a rigorous yet accessible source of original research of interest to anyone considering questions pertaining to the philosophy of big data, especially as it pertains to questions surrounding democracy, the democratic process and the role of modern forms of social media. These are clearly pressing issues facing contemporary democracies, and the chapters herein provide much-needed clarity in understanding and challenging those issues.

References


