



BRILL

PUBLIC ANTHROPOLOGIST 3 (2021) 119-147

PUBLIC  
ANTHROPOLOGIST

brill.com/puan

## Articles



# Could Cambridge Analytica Have Delivered Donald Trump's 2016 Presidential Victory? An Anthropologist's Look at Big Data and Political Campaigning

*Vito Laterza*

Associate Professor, Department of Global Development and Planning  
& Centre for Digital Transformation (CeDiT), University of Agder,  
Kristiansand, Norway  
[vito.laterza@uia.no](mailto:vito.laterza@uia.no)

### Abstract

I first provide some context about Cambridge Analytica's (CA) activities, linking them to CA parent company, SCL Group, which specialised in "public relations" campaigns around the world across multiple sectors (from politics to defence and development), with the explicit aim of behavioural change. I then analyse in more detail the claims made by mathematician and machine learning scholar David Sumpter, who dismisses the possibility that CA might have successfully deployed internet psychographics (e.g. online personality profiling) in the winning 2016 Trump presidential campaign in the US. I critique his arguments, pointing at the need to focus on the bigger picture and on the totality of CA methods, rather than analysing psychographics in isolation. This is followed by a section where I use CA whistleblower Christopher Wylie's 2019 memoir to show the important role that in-depth qualitative research and methods akin ethnographic immersion might have played in building CA big data capabilities. I provide an angle on big data that sees it as complementary, rather than in opposition to, human insight that comes from qualitative immersion in the social realities targeted

by CA. The concluding section discusses additional questions that should be explored to gain a deeper understanding of how big data is changing political campaigning, with an emphasis on the important contribution that anthropology can make to these crucial debates.

### Keywords

Cambridge Analytica – big data – political campaigns – psychometrics – internet psychographics – qualitative research – 2016 US presidential election – anthropology of algorithms

More than four years have passed since the political earthquake caused by Donald Trump's US presidential victory in 2016, and yet the debate over the causes and conditions that made such victory possible continues unabated. Was it a backlash against the rise of the first and only US president of colour Barack Obama? Did Trump manage to mobilise white working class voters with a mix of "economic nationalism" and anti-immigration rhetoric? Was the vote about economic anxiety or racial mobilisation? Whatever one thinks about these thorny questions, one discussion seems particularly hard to settle: could Trump have won, after all, because of data campaigning by the now defunct UK-based political consulting firm Cambridge Analytica (CA from now onwards)? Did big data and online personality profiling allegedly used by CA deliver his victory?

Trump has lost the November 2020 presidential contest to the Democratic Party's candidate Joe Biden. But the tycoon's support in actual votes rose by around 11 million since 2016, indicating that his 2020 campaign – also, according to many media investigations, driven by data campaigning and social media messaging – might have failed to deliver his re-election, but was nonetheless highly effective.<sup>1</sup> Trump's propaganda machinery was working against rather unfavourable material conditions. Yet, even after a disastrous handling of the pandemic which resulted in hundreds of thousands of deaths and massive

---

<sup>1</sup> Coppins, M. (2020). The Billion-Dollar Disinformation Campaign to Reelect The President. *The Atlantic*, 10 February; Ryan-Mosley, T. (2020). The Technology that Powers the 2020 Campaigns, Explained. *MIT Technology Review*, 28 September; Wong, J.C. (2020). One Year Inside Trump's Monumental Facebook Campaign. *The Guardian*, 29 January.

negative economic consequences for the majority of Americans, Trump's campaign was able to embolden his base and gain new voters, and to convince a majority of his supporters that the US response to the pandemic was going well and that the US economy was in good or excellent conditions.<sup>2</sup> At the local level, the 2020 Trump campaign reached the media spotlight in the aftermath of the election for having successfully targeted Florida's Latinx communities (many of Cuban and Venezuelan origin) with messaging depicting Biden as somebody who would bring Cuban or Venezuelan-style socialism to the US – Trump did win Florida, and Latinx support for him there increased significantly since 2016.<sup>3</sup>

We will likely see more academic studies of the 2020 Trump campaign in the coming months and years, but for now we do have already at our disposal a large and constantly increasing amount of evidence on the 2016 campaign and on the role played by CA. This information comes from journalistic investigations, whistleblowers' accounts, document leaks, and commissions of inquiries from parliaments and government agencies in the US and the UK. While of course we cannot unproblematically assume that the methods used by CA in 2016 are the same as those deployed by the Trump campaign in 2020, the 2016 events provide crucial background for subsequent campaigns.

Despite the available evidence, our academic understanding of these issues is in formation, and there is a relative dearth of in-depth social scientific analyses of the CA data scandal, especially more qualitatively oriented ones – and anthropology is no exception.<sup>4</sup>

2 See CNN exit polls for the US 2020 presidential election, <https://edition.cnn.com/election/2020/exit-polls/president/national-results> (Accessed 22 November 2020).

3 Agiesta, J., Luhby, T., Sparks, G., Struyk, R. (2020). More Latino Voters Support Trump in 2020 than 2016, but Young Americans Favor Biden, Early CNN Exit Polls Show. *CNN*, 5 November; Santiago, F. (2020). In Florida, Biden Couldn't Shake Trump's Lie That Democrats Are Radical Socialists. *Miami Herald*, 4 November; Taladrid, S. (2020). How Pro-Trump Disinformation Is Swaying A New Generation of Cuban-American Voters. *The New Yorker*, 26 October.

4 To my knowledge, there are barely any peer-reviewed pieces by anthropologists that have a substantial focus on CA. They include: González, R.J. (2017). Hacking the Citizenry? Personality Profiling, "Big Data" and the Election of Donald Trump. *Anthropology Today* 33(3): 9–12; Laterza, V. (2018). Cambridge Analytica, Independent Research and the National Interest. *Anthropology Today* 34(3): 1–2. While the broader social scientific literature has several articles that focus on some aspects of the CA data scandal, in-depth accounts focusing specifically on the company and its methods are few. See in particular: Berghel, H. (2018). Malice Domestic: The Cambridge Analytica Dystopia. *Computer* 51(5): 84–89; Richterich, A. (2018). How Data-Driven Research Fuelled the Cambridge Analytica Controversy. *Partecipazione e Conflitto* 11(2): 528–543; Risso, L. (2018). Harvesting Your Soul? Cambridge Analytica and Brexit. *The Selected Proceedings of the Symposium "Brexit Means Brexit,"* Akademie der Wissenschaften und der Literatur, Mainz; Ward, K. (2018). Social Networks, the 2016 US Presidential Election,

The available academic literature often eludes in-depth assessments of whether CA methods could have delivered Trump's election. Some of those who have taken a clearer stance in this matter suggest that CA methods could have indeed shifted voter behaviour in unethical, if not illegal, ways. These analyses are frequently informed by concerns about privacy and the integrity of the democratic process.<sup>5</sup>

In his analysis of the CA case, anthropologist Roberto J. González expresses scepticism that big data could have played a decisive role in delivering Trump's election.<sup>6</sup> In this, he joins a chorus of other scholars and public intellectuals who have dismissed that possibility.<sup>7</sup>

Scholars that are more worried about the possible influence of CA tactics on voter behaviour have not on the whole tried to debunk the arguments of the sceptics.<sup>8</sup> In this article, I will try to fill this gap by providing a detailed critique of some of the arguments put forward by sceptics. My work joins the growing scholarly literature that has been sounding alarm bells about CA and similar digital campaigning operations.<sup>9</sup>

---

and Kantian Ethics: Applying the Categorical Imperative to Cambridge Analytica's Behavioral Microtargeting. *Journal of Media Ethics* 33(3): 133–148.

The vast majority of the academic discussions focus on the case of the US 2016 presidential election campaign, and only a handful consider the work of CA in Africa, for example: Ekdale, E., Tully, M. (2019). African Elections as a Testing Ground: Comparing Coverage of Cambridge Analytica in Nigerian and Kenyan Newspapers. *African Journalism Studies*; Maweu, J.M. (2020). "Fake Elections?" Cyber Propaganda, Disinformation and the 2017 General Elections in Kenya. *African Journalism Studies*; Nyabola, N. (2018). *Digital Democracy, Analogue Politics: How the Internet Era is Transforming Politics in Kenya*. Zed Books. Even though the accounts focusing primarily on CA are relatively few, the influence of the CA data scandal on scholarship is significant when we consider the rapidly growing social scientific literature on online political campaigning and the role of big data, algorithms and microtargeting in influencing political behaviour. See for instance: Amakoh, K.O., Faustino, B.A., Oloruntoba, F.A. and Ogwezzy-Ndisika, A.O. (2018). Big Data And Accountability in Nigeria: Insights from the BudGIT Organization and the #OpenNASS Campaign. *Partecipazione e Conflitto* 11(2): 472–483; Bodó, B., Helberger, N. and de Vreese, C.H. (2017). Political Micro-Targeting: A Manchurian Candidate or Just a Dark Horse? *Internet Policy Review* 6(4); Hoferer, M., Böttcher, L., Herrmann, H.J. and Gersbach, H. (2020). The Impact of Technologies in Political Campaigns. *Physica A*, 538: 1–12.

5 For example: Berghel, Malice Domestic; Ward, Social Networks.

6 González, Hacking The Citizenry?

7 See for instance: Kavanagh, C. (2018). Why (Almost) Everything Reported about the Cambridge Analytica Facebook "Hacking" Controversy is Wrong. *Medium*, 25 March; Levine, Y. (2018). The Cambridge Analytica Con. *The Baffler*, 21 March; Sumpter, D. (2018). *Outnumbered: From Facebook and Google to Fake News and Filter-Bubbles. The Algorithms That Control Our Lives*. Bloomsbury.

8 One exception is Laterza, Cambridge Analytica.

9 See for example: Berghel, Malice Domestic; Nyamnjoh, F. (2018). *The Rational Consumer: Bad for Business and Politics – Democracy at the Crossroads of Nature and Culture*. Langaa RPCIG.

My approach to public anthropology is interdisciplinary. I weave anthropological and social scientific concerns, to show the limits of understanding complex cases such as the CA data scandal exclusively through disciplinary lenses. The contribution anthropology can make to these issues assumes a bifocal approach: a reflexive look at our own discipline and its conceptual and empirical tools, in parallel with an interdisciplinary dialogue that makes such insights relevant to a wider range of academic and non-academic audiences beyond the disciplinary boundaries of anthropology.

We will probably never be able to provide a definitive answer to the question of whether CA data capabilities delivered Trump's victory, mostly because we do not have the smoking gun, that is, the models and datasets used by CA are not available for scrutiny.<sup>10</sup> We can however productively assess the significant amount of evidence so far, especially if we reformulate the issue in the following way:

- is it possible that CA developed effective political campaigning tools, including, but not limited to, granular data-driven techniques, that could have given a significant advantage to Trump in the 2016 election?

Two recent books by CA whistleblowers published in October 2019 – Christopher Wylie's *Mindf\*ck* and Brittany Kaiser's *Targeted* – provide us with nuanced and layered testimonies of how CA worked from the inside.<sup>11</sup> The accounts move beyond, but also help us make sense of, the fragmented evidence from the various commissions of inquiries, leaks and journalistic investigations.

The fact that both authors were social science PhD candidates at the time of working for CA might help to explain why their views might be particularly relevant for a social scientific analysis of the whole affair. Wylie was doing a PhD

<sup>10</sup> A Channel 4 investigation released in September 2020 might constitute an important breakthrough in this direction: Channel 4 claims to have in their possession a Cambridge Analytica database with detailed personal information on almost 200 million American voters, that was allegedly used by the company as the basis for its microtargeting activities for the 2016 Trump campaign. The investigations show convincing evidence that triangulates the details of the database with voter data from specific areas where possible CA activities to increase Trump support or suppress support for Clinton might have taken place. It is hoped that Channel 4 will safely share its database with academic researchers so that more studies can be conducted to systematically investigate the potential effects of CA activities on the US 2016 election results. See: Channel 4 News Investigations Team (2020). Revealed: Trump Campaign Strategy to Deter Millions of Black Americans from Voting in 2016. *Channel 4*, 28 September; Guru-Murthy, K. (2020). How Trump Campaign Targets Millions of White Voters – and Activates Fears Over Rioting. *Channel 4*, 29 September.

<sup>11</sup> Kaiser, B. (2019). *Targeted: My Inside Story of How Big Data, Trump, and Facebook Broke Democracy and How It Can Happen Again*. HarperCollins; Wylie, C. (2019). *Mindf\*ck: Cambridge Analytica and the Plot to Break America*. Random House.

in machine learning and fashion at Central Saint Martins in London when he was hired by SCL Group, CA parent company, at the end of 2013.<sup>12,13</sup> In his book, he highlights the connections between his interests in fashion and politics, as he was keen to develop big data models that explored how aspects of social and cultural identity such as fashion styles are related to political attitudes and behaviour.<sup>14</sup> In March 2018, his whistleblowing activities became public, and in January 2019, true to his passion, Wylie took up the post of research director of Swedish fashion retailer H&M.<sup>15</sup> There was also a pragmatic reason that seemed to make employment with SCL attractive: SCL agreed to pay the tuition fees for his PhD, which were substantial, given his status as a Canadian international student in the UK.<sup>16</sup>

Financial considerations seemed to weigh big on Kaiser's decision to join SCL too. When she was offered a job by SCL director Alexander Nix in October 2014, she was in need of money due to an ongoing financial crisis in her family. Kaiser claims that it was her economic situation that made her set aside her reservations about working for a company involved with US Republican campaigning<sup>17</sup> – she describes herself as a “lifelong Democrat and a devoted activist who had worked for years in support of progressive causes.”<sup>18</sup> Like Wylie, Kaiser was interested in the relationship between big data and politics, and she juggled her consultancy activities for SCL and CA with PhD studies focused on “preventive diplomacy,” which included a focus on how big data could inform international peacekeeping organisations to prevent conflict in high-risk areas around the world.<sup>19</sup>

Thanks to these whistleblowers' accounts, we can now get a better sense of how CA worked as a coherent, complex organisation made first and foremost

12 Wylie, *Mindf\*ck*, p. 47.

13 CA was incorporated in January 2015. SCL (Strategic Communication Laboratories) Group had been around from much longer and sprang from its predecessor (and later partner), the Behavioural Dynamics Institute, which was founded by Nigel Oakes in 1990. Multiple sources report the birth of SCL (later SCL Group) in 1993; see: Meyer, J. (2018). Cambridge Analytica Boss Went From “Aromatics” to Psyops to Trump’s Campaign. *Politico*, 22 March; Pomerantsev, P. (2019). *This is Not Propaganda: Adventures in the War Against Reality*. Faber and Faber, p. 232. The UK company registry (Companies House) reports that SCL Group was incorporated in 2005. CA and SCL were so closely entangled that the boundaries between the two were often blurred.

14 Wylie, *Mindf\*ck*, pp. 45–47.

15 Harper, L. (2019). Whistleblower Christopher Wylie Joins Fashion Retailer H&M. *The Guardian*, 31 January.

16 Wylie, *Mindf\*ck*, p. 47.

17 Kaiser, *Targeted*, pp. 30–31.

18 *Ibid.*, p. 2.

19 *Ibid.*, pp. 16–17.

by humans, in a way that video bites, short media pieces and parliamentary inquiry reports cannot quite do. The books have an ethnographic feel that can push anthropological analysis further. While I use both works as sources to inform my analysis, I focus in this article primarily on Wylie's account, as he discusses the research methodologies employed by CA from the perspective of a qualitatively oriented data scientist. Kaiser's book is more focused on the business and marketing aspects – she worked with CA business development operations.<sup>20</sup>

I first provide some context about CA activities, linking them to CA parent company, SCL Group, which specialised in “public relations” campaigns around the world across multiple sectors (from politics to defence and development), with the explicit aim of behavioural change. I then analyse in more detail the claims made by mathematician and machine learning scholar David Sumpter. In a recent book, Sumpter dismisses the possibility that CA might have successfully deployed internet psychographics (e.g. personality profiling from online user data) in the 2016 Trump campaign.<sup>21</sup> I critique his arguments, pointing at the need to focus on the bigger picture and on the totality of CA methods, rather than analysing psychographics in isolation. This is followed by a section where I use CA whistleblower Christopher Wylie's memoir to show the important role that in-depth qualitative research might have played in building CA big data capabilities.<sup>22</sup> I provide an angle on big data that sees it as complementary, rather than in opposition to, human insight that comes from qualitative immersion in the social realities targeted by CA. The concluding section discusses additional questions that should be explored to gain a deeper understanding of how big data is changing political campaigning, with an emphasis on the important contribution that anthropology can make to these crucial debates.

The analysis is informed by insights from critical propaganda studies and from work on dual use anthropology.<sup>23</sup> The CA story is not just one of effective

20 The two memoirs complement each other temporally as well. Wylie joined SCL at the end of 2013 and left in the second half of 2014; Kaiser joined SCL at the end 2014 and left CA in early 2018; Wylie, *Mindf\*ck*, pp. 46–47, 144–154; Kaiser, *Targeted*, pp. 31–37, 309–310.

21 Sumpter, *Outnumbered*.

22 Wylie, *Mindf\*ck*.

23 An important text in critical propaganda studies that has influenced this work is: Briant, E.L. (2015). *Propaganda and Counter-Terrorism: Strategies for Global Change*. Manchester University Press. On dual use anthropology, see González, R.J. (2009). *American Counterinsurgency: Human Science and the Human Terrain*. Prickly Paradigm Press; Price, D.H. (2016). *Cold War Anthropology: The CIA, the Pentagon, and the Growth of Dual Use Anthropology*. Duke University Press.

(or ineffective) aggressive tactics in voter manipulation and of spectacular breaches of citizens' privacy. It is also one of many instances where scientific research and academic personnel have been deployed from the beginning as *the* weapon for unethical and often illegal propaganda activities, ranging from foreign interference in domestic politics of Northern and Southern countries, to working in military and security operations directed by powerful and unaccountable government and corporate interests.

Together with González, my article contributes towards filling an important gap in the nascent anthropology of algorithms: the study of the use of algorithms in political campaigns.<sup>24</sup> It provides one possible answer to Nick Seaver's question "What should an anthropology of algorithms do?": it should critically explore, among other things, the use and abuse of algorithms in election campaigning and political communication more broadly.<sup>25</sup>

Besteman and Gusterson have edited an important collection advancing the state of the art in the discipline.<sup>26</sup> The style of the collection is accessible and, for the most part, keeps explicit scholarly citations to a substantial set of endnotes. There is generally little canon anthropology mentioned in the chapters, except for the introduction by Gusterson, who acknowledges that the critique of algorithms has, so far, not seen anthropologists in a prominent position – which is the reason why most of the academic literature explicitly discussed by Gusterson in his introduction, by Besteman in her afterword, and by their colleagues in other chapters, is not anthropological.<sup>27</sup> While some chapters are based on ethnographic data, several others (such as one on ubiquitous surveillance and another on the logics of quantification) are not, showing that, to study what many now refer to as the "black box" of algorithms, we are frequently pushed beyond ethnography.<sup>28</sup> In the introduction to a seminal short collection of essays on the anthropology of big data, Boellstorff and

24 Besteman, C., Gusterson, H., eds. (2019). *Life by Algorithms: How Roboprocesses Are Remaking Our World*. University of Chicago Press; Boellstorff, T., Maurer, B. eds. (2015). *Data, Now Bigger And Better! Prickly Paradigm Press*; Seaver, N. (2018). What Should An Anthropology of Algorithms Do? *Cultural Anthropology* 33(3): 375–385.

25 Seaver, What Should An Anthropology.

26 Besteman, Gusterson, *Life by Algorithms*.

27 Gusterson, H. (2019). Introduction. In: Besteman, Gusterson, *Life by Algorithms*, pp. 1–27; Besteman, C. (2019). Afterword. In: Besteman, Gusterson, *Life by Algorithms*, pp. 165–180.

28 Masco, J. (2019). Ubiquitous Surveillance. In: Besteman, Gusterson, *Life By Algorithms*, pp. 125–144; Merry, S. (2019). Controlling Numbers: How Quantification Shapes the World. In: Besteman, Gusterson, *Life by Algorithms*, pp. 145–163; Gusterson, Introduction, p. 19.



Maurer make a similar point, and mention Tim Ingold's influential intervention "Anthropology is not ethnography."<sup>29</sup>

In terms of its ethnographic contribution, Besteman and Gusterson's edited collection provides an important range of empirical cases from the US which focus on the use of algorithms in home foreclosures, standardised testing and evaluation in education, detention and deportation of minors in immigration custody, and felony convictions.<sup>30</sup> These ethnographic accounts are centred around the notion of "roboprocess," which, in Gusterson's theorisation, emphasises the interaction between people and algorithms as an automated algorithmic process that effectively traps humans in specific situations where the "common sense and situational logic of humans is displaced by and subordinated to the logic of automation and bureaucracy."<sup>31</sup>

The primary focus of most chapters is on the interface between humans and technology in people's daily encounters with algorithms: the people that the authors engage with tend to be on the receiving end of the negative and often unintended effects of the algorithm, or include the middle-ranking clerks and officials executing the algorithms' decisions.

My article fills an important gap in this respect: the focus in my discussion is on the methods and techniques that those who create and manipulate algorithms employ, and on the effects of the algorithms seen from the perspective of those who are in the control room so to speak – in this I follow Seaver's call for studying the human in and behind the algorithms, and for avoiding the fetishisation of a supposed "autonomy" of algorithms "freed" from human factors.<sup>32</sup> While my analysis is not ethnographic – I did not have direct access to the main players of the Cambridge Analytica data scandal and their activities – it remains nonetheless anthropological in style and scope. The anthropology envisaged here is one that is not tied to a particular canon, and in line with Besteman and Gusterson, aims to develop an anthropological sensibility

29 Boellstorff, T., Maurer, B. (2015). Introduction. In: Boellstorff, T., Maurer, B., *Data, Now*, pp. 1–6; Ingold, T. (2008). Anthropology Is Not Ethnography. *Proceedings of the British Academy* 154: 69–92.

30 Stout, N. (2019). Automated Expulsion in the U.S. Foreclosure Epidemic. In: Besteman, Gusterson, *Life by Algorithms*, pp. 31–43; Lutz Fernandez, A., Lutz, C. (2019). Roboeducation. In: Besteman, Gusterson, *Life by Algorithms*, pp. 44–58; Terrio, S.J. (2019). Detention and Deportation of Minors in U.S. Immigration Custody. In: Besteman, Gusterson, *Life by Algorithms*, pp. 59–76; Middlemass, K.M. (2019). A Felony Conviction as a Roboprocess. In: Besteman, Gusterson, *Life by Algorithms*, pp. 77–87.

31 Gusterson, Introduction, p. 2.

32 Seaver, What Should an Anthropology, p. 378.

directed to a wider audience which includes both anthropologists and other social scientists interested in these topics.<sup>33</sup>

I have followed the Cambridge Analytica scandal and the 2016 US presidential election from early on, and my insight is coupled by a long-term engagement as a political commentator with the rise of social media politics, spontaneous protests, and the social media messages of political movements across the political spectrum in the US and Italy.<sup>34</sup> In recent years, I have also participated as an activist scholar in various transnational workshops, conferences and informal conversations with progressive party and social movement activists from around the world, and especially from Africa, Europe and North America. In all these events, discussions about the uses and abuses of social media for political campaigning were key, and I had the privilege of rapidly updating my knowledge with expert insiders, including social media campaigners and cybersecurity privacy activists. It is difficult to call this knowledge “ethnographic” in the classic sense, but there is no doubt that I am speaking to some extent “from the inside,” and have been autoethnographically immersed in these flows and events – and the “hybrid networks” that they encompass – since my early days as a direct action activist in Cambridge in 2009–2010.<sup>35</sup>

### What Did Cambridge Analytica Actually Do?

Cambridge Analytica and its parent company SCL Group shut down in May 2018. CA was hired by Trump’s successful 2016 presidential campaign, and is alleged to have played an important part in the Leave campaigns of the Brexit referendum – although its role in the latter remains unclear.<sup>36</sup> The firm was tightly knit with, and often indistinguishable in practice from, its parent company SCL Group. Both companies were involved in several more election

33 Besteman, Gusterson, *Life by Algorithms*.

34 See for example: Laterza, V. (2012). Innocence of Muslims: How Fiction Creates Reality. *Al Jazeera English*, 19 September; Laterza, V. (2016). Democracy After Sanders: Building A Progressive Alternative, Beyond Social Media And Mass Rallies. *openDemocracy*, 26 May; Laterza V. and Romer, L.P. (2020). With the Support of the Left, Biden Can Deliver Progressive Gains. *Al Jazeera English*, 21 November.

35 The term “hybrid networks” is borrowed from: Castells, M. (2015). *Networks of Outrage and Hope: Social Movements in the Internet Age*. 2nd edition. Polity Press.

36 Hern, A. (2019). Cambridge Analytica Did Work for Leave.EU, Emails Confirm. *The Guardian*, 30 July; Lucas, I. (2020). New Evidence Connects Johnson, Gove and Cummings to Cambridge Analytica. *Byline Times*, 23 October.

campaigns around the world, including work in Indonesia, Nigeria and Kenya, all cases widely covered by the media in the aftermath of the CA data scandal.<sup>37</sup>

CA was born as a partnership between SCL Group (also UK-based), US far right propagandist and former Breitbart editor Steve Bannon, US right-wing billionaire and technologist Robert Mercer, and Robert's daughter Rebekah Mercer. British director Alexander Nix was the key person from SCL who line managed CA data scientists and actively procured contracts.<sup>38</sup> Even though CA was incorporated in 2015, its protagonists were already working together for several Republican candidates in the US 2014 mid-term elections.<sup>39</sup> CA joined the Trump campaign in June 2016.<sup>40</sup> Before that, the company had worked for Ted Cruz's 2016 presidential primary campaign.

SCL Group ran four divisions: SCL Defence (running defence contracts primarily with US and UK state agencies), SCL Social (running projects funded from development aid), SCL Elections (the division most closely tied to CA activities, but predating CA by several years) and SCL Commercial (running contracts with the private sector, which CA also did).<sup>41</sup> Despite the broad variety of sectors the two companies engaged with, the common thread was that they specialised in "public relations" campaigns aimed at behavioural change – the latter is a label well known to anthropologists of development, think for instance about behavioural change campaigns designed to increase condom use in HIV-affected countries.

SCL expertise had developed on the back of the British think tank Behavioural Dynamics Institute (BDI), founded in 1990 by Nigel Oakes.<sup>42</sup> In 2010, the BDI presented itself as an "academic institute that specialises in understanding influence and persuasion in order to change audiences' attitudes and

37 Ghoshal, D. (2018). Mapped: The Breathtaking Global Reach of Cambridge Analytica's Parent Company. *Quartz*, 28 March.

38 Kaiser, *Targeted*; Wylie, *Mindf\*ck*.

39 Wylie, *Mindf\*ck*.

40 Iling, S. (2018). Cambridge Analytica, The Shady Data Firm That Might Be A Key Trump-Russia Link, Explained. *Vox*, 4 April.

41 For an academic account of some of SCL defence propaganda activities, see Briant, *Propaganda*. Emma Briant also submitted three essays as evidence to the UK House of Commons, accompanied by recordings of some of the field interviews behind her 2015 book. These essays are a key source of reliable data and analysis about CA and SCL activities; Briant, E.L. (n.d.). Three Explanatory Essays Giving Context And Analysis To Submitted Evidence. Submitted to the Digital, Culture, Media and Sport Committee of the UK House of Commons, <https://old.parliament.uk/business/committees/committees-a-z/commons-select/digital-culture-media-and-sport-committee/news/fake-news-briant-evidence-17-19/> (Accessed 29 October 2020).

42 Pomerantsev, *This Is Not Propaganda*, pp. 229–235.

behaviour. The institute specialises in applying its methodology to military and political campaigns, where the audiences are hostile or friendly, national or international.”<sup>43</sup> The BDI “methodology draws extensively from group and social psychology and incorporates semiotics, semantics and many elements of cultural anthropology.”<sup>44</sup> The BDI stated goals hinted at ambitions that went well beyond academic research, and in practice the distinction between BDI and SCL/CA was blurred, with several personnel and projects straddling across the academic and business operations.

Named academic members and contributors on their archived website ranged across many disciplines, including clinical and social psychology, sociology, political science, international relations, and media & communications. Many were based in academic social science departments in allegedly “independent” units in well renowned universities, while others had overt ties with military and security agencies in the UK and the US. Some had strong affiliations with both worlds. Anthropology featured in some of the academic interests listed, but none of the people profiled with their bios in the BDI website seem to have had graduate training in anthropology. However, there are indications that anthropology might have played a more direct role, as journalist and disinformation expert Peter Pomerantsev claims in his 2019 book on social media and propaganda:

Oakes pioneered surveys by teams of anthropology students, who, usually without revealing their mission, spent long periods penetrating a community, enquiring about who people hated and trusted, what they most desired, which friends would influence them, what dictated how they behaved within a group.<sup>45</sup>

Pomerantsev’s account of the pre-social media days of SCL seem in fact to give quite a lot of weight to the influence of long-term fieldwork on the development of SCL research techniques.<sup>46</sup> And of course the countries where SCL worked in political, and later military and development campaigns had long been, at least in the Anglo-American academe, the remit of anthropological knowledge.

43 [https://web.archive.org/web/20100403074355/http://www.bdinstitute.org/03\\_what.html](https://web.archive.org/web/20100403074355/http://www.bdinstitute.org/03_what.html) (Accessed 29 October 2020).

44 [https://web.archive.org/web/20100403074355/http://www.bdinstitute.org/03\\_what.html](https://web.archive.org/web/20100403074355/http://www.bdinstitute.org/03_what.html) (Accessed 29 October 2020).

45 Pomerantsev, *This Is Not Propaganda*, p. 231.

46 *Ibid.*, pp. 231–233.

In its more recent data-driven incarnation, CA itself deployed several data scientists and scientific knowledge from academia, in particular from the University of Cambridge. The choice of the name was not simply an act of “prestige appropriation,” it also indicates an active recruitment of Cambridge data scientists and deployment of cutting edge research on internet psychographics developed at the university.<sup>47</sup> According to leaked documents, CA used the BDI methodology, centred around Target Audience Analysis (TAA), a method that is well established in Anglo-American military propaganda circles and has been discussed in many defence publications.<sup>48</sup>

The alleged innovation of this method in the propaganda and psychological operations (PSYOP) field is that it aims to study individual, social, cultural, political and economic characteristics of specific groups and their members before a specific campaign is launched. This then informs the crafting of messages that are directed at such groups or some segments of them, and that build on the knowledge gained about a combination of various psychological and socio-cultural traits emerging from the analysis.

Messages are tailored to their target audiences, and knowing such audiences becomes a key priority for successful behavioural change campaigns. The original insights of TAA were further developed by CA, which coupled them with the power of big data – including, but not limited to, personality profiling of social media users – with the aim of significantly increasing the effectiveness and reach of targeted messages.<sup>49</sup>

### Internet Psychographics and Beyond

A whole host of arguments have been put forward by academics and public intellectuals to the effect of dismissing the possibility that CA might have in fact delivered Trump’s victory. One common element in many of these analyses has been to focus on CA claim that it used internet psychographics to develop successful targeted messages that would bring a candidate to victory.

Many scholars argued that it is unlikely that this was the case, and some highlighted that this was probably just a hyped business message to promote

---

47 Wylie, *Mindf\*ck*, pp. 95–111.

48 See for instance: Mackay, A. and Tatham, S. (2011). *Behavioural Conflict: Why Understanding People and Their Motivations Will Prove Decisive in Future Conflict*. Books Express Publishing; Tatham, S., Le Page, R. (2014). *NATO Strategic Communication: More to Be Done?* National Defence Academy of Latvia.

49 Briant, Three Explanatory Essays; Pomerantsev, *This Is Not Propaganda*.

the company, and that buying into that would in itself help the company get more clients.<sup>50</sup> Others pointed out that the moral panic around the use and abuse of personal data from online use in the CA case was somewhat misleading, as microtargeting and the use of big data to influence people's behaviour has been a staple of commercial companies and the tech sector for quite some time.<sup>51</sup>

Few among the sceptics however stress the close links between the military and security establishment and CA, and the history of applied social scientific propaganda studies carried out by SCL.<sup>52</sup> On the whole, critics tend to narrowly focus on internet psychographics, no doubt driven by legitimate doubts about some of CA marketing spin possibly overstating their case.<sup>53</sup> Many scholars who take more seriously the possibility that CA methods might have been effective also focus on internet psychographics and big data.<sup>54</sup>

For the purposes of this article, I would like to consider the arguments put forward by David Sumpter, a mathematician who works with modelling social behaviour and machine learning and published the popular nonfiction book *Outnumbered*. Sumpter's work provides a compelling and informative explanation of the complex statistical techniques behind internet algorithms, discussing a vast array of cases, including CA personality profiling claims.<sup>55</sup> I argue that Sumpter's analysis – and some important contradictions in it – provide key elements to interpret CA whistleblowers' accounts. By translating complex technical matters for a non-specialist audience, Sumpter helps us humanise the CA story, and evaluate the scientific logics at work in the use and manipulation of statistical data in algorithms and of what is commonly lumped in the category of big data.

I will critique his claim that CA could not actually have done much with psychographics, if they ever used them at all. At the same time, I will use some of Sumpter's arguments about social statistics and big data to show why CA could have indeed gone quite far in successfully influencing voter behaviour to Trump's advantage. It should be noted that Sumpter published his book in June 2018, more than a year before the whistleblowers' accounts and the most recent company document leaks by one of the whistleblowers, so it is quite

50 Lynch, G., Willis, J., Cheeseman, N. (2018). Claims about Cambridge Analytica's Role in Africa Should Be Taken with a Pinch of Salt. *The Conversation*, 23 March.

51 Kavanagh, Why (Almost) Everything; Levine, The Cambridge Analytica Con.

52 A notable exception is González, Hacking The Citizenry?

53 Sumpter, *Outnumbered*, p. 43.

54 See for instance: Isaak, J. and Hanna, M.J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer*, 51(8): 56–59.

55 Sumpter, *Outnumbered*.

possible that the mathematician's scepticism about CA data capabilities was informed by the evidence that was available to him then.<sup>56</sup>

Both Christopher Wylie and Brittany Kaiser have no doubt that a massive amount of data was collected on the US population. Their estimates range from detailed profiles of “tens of millions of Americans ..., with potentially hundreds of millions more to come”<sup>57</sup> to “some 5,000 data points on every single American over the age of eighteen.”<sup>58,59</sup> These data did not only come from Facebook, but also from the national census, credit histories and anything else CA could get its hands on. The exact amount and nature of the data collected remains unclear, and, according to Kaiser, the company would not go into details over what datasets it owned and how it obtained them.<sup>60</sup>

Building on rapidly growing research on the usage of social media data to determine personality traits, and in particular work carried out by Michal Kosinski, David Stillwell and other colleagues, Cambridge Analytica allegedly used the OCEAN personality model (Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism) to profile “large numbers of US voters.”<sup>61,62</sup> Wylie claims that research from the Department of Psychology

---

56 In January and in October 2020, Brittany Kaiser leaked a substantial number of CA company documents through the Twitter account @HindsightFiles. The files leaked in January 2020 are not available through the original links provided in the Twitter account anymore, but most of them have been stored by the Organized Crime and Corruption Reporting Project here: <https://aleph.occrp.org/datasets/1519> (Accessed 24 October 2020). The documents leaked in October 2020 are available here: <https://archive.org/details/ca-docs-with-redactions-sept-23-2020-4pm/page/n12/mode/2up> (Accessed 25 October 2020).

57 Wylie, *Mindf\*ck*, p. 111.

58 Kaiser, *Targeted*, p. 78.

59 It should be noted that the difference in magnitude here is also temporal: Wylie is referring to an earlier stage of CA research, while Kaiser is talking about the dataset in the later years of operations, likely to have been updated and expanded well after Wylie's exit from the company in the second half of 2014.

60 Kaiser, *Targeted*.

61 Sumpter, *Outnumbered*, p.43.

62 On the use of the OCEAN model for the analysis of online data, see for instance: Kosinski, M., Stillwell, D. and Graepel, T. (2013). Private Traits and Attributes Are Predictable From Digital Records of Human Behavior. *Proc Natl Acad Sci USA* 110(15): 5802–5805; Matz, S.C., Kosinski, M., Nave, G. and Stillwell, D.J. (2017). Psychological Targeting as an Effective Approach to Digital Mass Persuasion. *Proc Natl Acad Sci USA* 114(48): 12714–12719. The OCEAN model (also known as the Five-Factor model or the Big Five model) has its roots in the long history of personality traits assessment, a subfield of psychometrics (the science of psychological measurement and assessment). According to Digman, earlier versions of the Five-Factor model surfaced in the 1930s, but the framework only moved beyond niche status in the 1980s. See: Digman, J.M. (2002). Historical Antecedents of the Five-Factor

and the Psychometrics Centre at the University of Cambridge was key for the development of CA psychological targeting tools.<sup>63,64</sup>

According to widely popularised research carried out by Youyou, Kosinski and Stillwell, computer models based on Facebook “likes” were better than friends, family, spouses and co-workers at assessing somebody’s personality traits.<sup>65</sup> Among other things, CA allegedly developed campaign messages for their American clients building on this body of research, based on the assumption that such messages were able to change the target’s voting behaviour.<sup>66</sup>

Put it this way, it does sound like a pretty bold claim, especially for a model that is discussed in the literature in rather crude terms. In a meta-analysis of research on personality assessment from social media data, Azucar, Marengo and Settanni state that: “individuals with high openness tend to have larger networks;” “[i]ndividuals with high conscientiousness appear to be cautious in managing their social media profiles;” “individuals with high extraversion have been characterized by higher levels of activity on social media ... and have a greater number of friends ... than introverted individuals;” and “[i]ndividuals with high neuroticism ... use more negative words in their posts [while] agreeable individuals tend to use fewer swear words and express positive emotions more frequently in their posts.”<sup>67</sup>

From an anthropological perspective, OCEAN looks like little more than a simplistic straightjacket that surely cannot capture and describe the nuances and complexities of individual behaviour in real life social settings.

Simplistic or not, it is still worth taking a look at the available evidence: how was OCEAN actually used to craft campaign messages?

There is some evidence on the 2016 Trump campaign: in the pile of documents leaked by Kaiser in October 2020, a CA internal report dated November 2016 analyses CA activities during the campaign and their outcomes after the victory. As with other CA company documents, reading this report gives an eerie feeling, as it discusses in professional and technical language the effectiveness

---

Model. In: Costa, P.T. and Widiger, T.A. eds. *Personality Disorders and the Five-Factor Model of Personality*. 2nd edition. American Psychological Association, pp. 17–22.

63 Wylie, *Mindf\*ck*, pp. 95–96.

64 Kosinski and Stillwell have worked or continue to work at the Psychometrics Centre, and Kosinski did his PhD in Psychology at Cambridge.

65 Youyou, W. Kosinski, M. and Stillwell, D. (2015). Computer-Based Personality Judgements Are More Accurate Than Those Made by Humans. *PNAS* 112(4): 1036–1040.

66 Wylie, *Mindf\*ck*, pp. 103–109.

67 Azucar, D., Marengo, D. and Settanni, M. (2018). Predicting the Big 5 Personality Traits From Digital Footprints on Social Media: A Meta-Analysis. *Personality and Individual Differences*, 124, p. 151.



of campaign messages based on insults, smears and fake news. For instance, the project CA was tasked with by the Make America Number One Super PAC (primarily funded by Robert Mercer) is referred to in the internal report with the rather explicit label “Defeat Crooked Hillary”, then uncannily shortened, in classic business fashion, to DCH.<sup>68</sup> In the same vein, the report boasts that “[t]he entire [Make America Number One] team should take confidence in the knowledge that we did work other groups and individuals were unwilling to do in defeating Hillary Clinton.”<sup>69</sup>

The main focus of the project was to target around 9 million voters nationwide, with “a special focus on New Hampshire, Pennsylvania, Virginia, North Carolina, Florida, Ohio, Iowa, Colorado, Nevada, and Michigan,” battleground states which could have gone in either direction (e.g. Republican or Democrat) in the election.<sup>70</sup> The details provided on the target audience are sketchy, but the report suggests that the bulk of the voters targeted by CA digital media campaigns included Clinton voters who could be persuaded either not to vote for Clinton, or to vote for Trump.<sup>71</sup> Elsewhere in the same document, however, it is suggested that Republican voters were also targeted in “get out the vote” campaigns aimed at increasing their chances of showing up to vote.<sup>72</sup> The report claims that the most successful ads were run on Facebook and Google Search, but the full range of platforms used included YouTube, Twitter, Google Display Network, SnapChat, Pandora’s internet radio, email and traditional television.<sup>73</sup>

Psychographics are discussed towards the end, in a section on email advertising. Here it is claimed that CA tested its OCEAN psychographics with two email campaigns.<sup>74</sup> One email had a subject line that “was designed to be reassuring to people who ordinarily might have a propensity to worry” – it read “Preserve Freedom and Overcome Hillary’s Candidacy” and was crafted with voters scoring high on neuroticism in mind. The email was sent to a group of people who scored high on neuroticism and another group who did not. The first group had 20% higher open rates than the second.

68 Cambridge Analytica – Select 2016 Related Documents (added online on 28 September 2020), <https://archive.org/details/ca-docs-with-redactions-sept-23-2020-4pm/page/n77/mode/2up> (Accessed 29 October 2020), p. 76.

69 *Ibid.*, p. 90.

70 *Ibid.*, p. 76.

71 *Ibid.*, p. 76.

72 *Ibid.*, p. 77, 84. The targeting of different segment of Clinton and Trump voters described in this report is aligned with a slide of a CA presentation in the same pile of leaked documents; *ibid.*, p. 99.

73 *Ibid.*, pp. 78–79, 89–90.

74 *Ibid.*, pp. 88–89.

The second email campaign targeted only people scoring high on neuroticism, and had three types of email subject – one with a reassuring message (for example, “Calm the storm, stop Hillary”), another one with a fear-driven message (for example, “Electing Hillary destroys our nation”), and a last one with a generic message (for example, “Information from Make America Number 1”). The fear-based subject line had 10% higher open rate than the generic one, and 20% more opens than the reassuring one.

The report concludes that “[t]hese email campaigns demonstrate the effectiveness of psychographic profiling for enhancing email marketing campaigns.”<sup>75</sup>

Kaiser’s leaks also contain documents on the John Bolton Super PAC, which funded victorious Republican candidates in the 2014 US mid-term elections in Arkansas and North Carolina.<sup>76</sup> This set of documents provides important evidence about how CA allegedly approached the use of psychographics. In one example from North Carolina, the company documents discuss two groups that were targeted with different messages about the same issues:

North Carolina Group 3 consisted of young, female voters who displayed high neuroticism and cared most about the economy, national security and immigration. These voters were shown advertisements that highlighted the failures of the current administration’s national security policy.

North Carolina Group 4 consisted of an even split of male and female voters who displayed high conscientiousness and agreeableness. These voters cared most about the economy and education, so were shown advertisements that positioned national security as a family and social issue.<sup>77</sup>

Perhaps there is nothing here as mind-blowing as some sensationalistic accounts seem to imply, but nor is there anything that suggests outright that such an approach did not or could not work. At the very least, this evidence

<sup>75</sup> Ibid., p. 89.

<sup>76</sup> You can find the link to the John Bolton Super PAC files leaked in January 2020 here: <https://aleph.occrp.org/entities/43385707.074609bfc8dffib3c9c63236c811d9964fdc3d05> (Accessed 24 October 2020). There is also a substantial set of additional documents on the use of psychographics in the same work for John Bolton Super PAC in the October 2020 leaks; Cambridge Analytica – Select 2016 Related Documents.

<sup>77</sup> The source document is available here: <https://aleph.occrp.org/entities/d81dc9c2d6096cf748ad735624e3e181e9c93334.2c4a3fa5e945bf6ed82f2e580df940052f8b05ff#page=1> (Accessed 24 October 2020).

should push us to inquire further about the effectiveness of microtargeting in election campaigning.

From the perspective of an expert modeller and statistician versed in such algorithmic methods, Sumpter tells us as much: computer models of human behaviour are often too simplistic to have high predictive validity, and when you compare humans with machine learning algorithms, usually humans match or beat the algorithms' predictions – he provides a vast array of examples to make this point.<sup>78</sup>

When it comes to CA, Sumpter tested the Facebook-based OCEAN model of Michal Kosinski and colleagues with an anonymised dataset of 20,000 users they released for psychology students.<sup>79</sup> For four of the five OCEAN traits, the Facebook-based model recreated by Sumpter correctly predicted those traits in 60% of the cases – only slightly better than leaving it to chance (50%). For one trait (openness), the model predicted correctly two thirds of the cases. Based on these tests, Sumpter's conclusion is that personality assessment does not work – and hence would have been unlikely to work in the case of CA as well.

There are however a number of other factors emerging from the available evidence, and including claims Sumpter himself makes, that run counter such a clear cut conclusion. First, both Wylie's and Kaiser's accounts show that CA data operations were much more than just psychographics: with allegedly massive amounts of data about possibly hundreds of millions of Americans, OCEAN was just one tool in the data arsenal of the firm.<sup>80</sup>

Audience segmentation of voters carried out by CA was not just based on OCEAN, but included geolocation, census, health, credit and myriads more personal data that CA could get hold of. Both accounts also clearly say that internet psychographics were involved in only some of the models they used: the general point with algorithms – as Sumpter himself shows throughout his book – is that they can find out all kinds of patterns and statistical correlations from user data that would otherwise not be known with traditional statistical and qualitative research methods.

This means that algorithms can find valid patterns that humans who work with them, or other social scientists, might have no explanation for. Unlike earlier forms of artificial intelligence that tried to reproduce certain human

---

78 Sumpter, *Outnumbered*.

79 *Ibid.*, pp. 50–54.

80 Wylie, *Mind\*ck*; Kaiser, *Targeted*.

logics, algorithms behind contemporary machine learning work on probability and do not need a “theory” to work.<sup>81</sup>

In some instances the colour of a button in political Facebook ads can be a decisive factor in clickthrough rates (how many times people click on the ad), and hence live experiments with the ads allow the online campaign team to make quick and continuous adjustments to improve the ads’ effectiveness.<sup>82</sup> Why do people with certain characteristics extracted from algorithmic analysis react better to one colour than another? That could be a whole research project, but fundamentally *why* that is so is irrelevant to *how* effective it is.

This also means that models based on the integration of OCEAN with other data and assumptions, can potentially improve the efficacy of the political messages, regardless of how sound the assumptions are from the perspective of social scientific knowledge.

The descriptions of the CA virtual dashboards available to campaigners suggest something akin NSA-like real time in-depth surveillance that can inform anything from physical campaigners knocking on somebody’s door already knowing very detailed personal information about the house residents, to real time automatic adjustments of the kind and quantity of political ads, and who they are targeted to, through monitoring sophisticated feedback statistics about their effects.<sup>83</sup> The internal report on the effectiveness of CA tactics for the 2016 Trump campaign mentioned before throws light on what looks like a fine-tuned, relatively complex but also quite clear and well planned digital methodology that supports and scales up human insight and decision-making beyond what human agency on its own could have achieved, with a relatively small and focused amount of resources. When it comes to money, we need to take CA claims with even more caution than usual, but the CA internal report states that they received around USD 5.6 mil from Make America Number One Super PAC – a relatively small amount for the reach the campaign claimed to have had.

---

81 Boellstorff makes a similar point when he notes that “[a]lgorithmic living is displacing artificial intelligence as the modality by which computing is seen to shape society: a paradigm of semantics, of understanding, is becoming a paradigm of pragmatics, of search. Contemporary computational language translation, for instance, does not work by trying to get a computer to intelligently understand language: systems like Google Translate work by matching texts from a vast corpus, without the computer ever ‘knowing’ what is said”; Boellstorff, T. (2015). Making Big Data, in Theory. In: Boellstorff, T. and Maurer, B., *Data, Now*, pp. 90–91.

82 Denning, S. (2018). How Agile Helped Elect Donald Trump. *Forbes*, 22 July; Loizos, C. (2017). “When You Spend \$100 Million on Social Media,” It Comes With Help, says Trump strategist. *TechCrunch*, 8 November.

83 Kaiser, *Targeted*, pp. 87–88.

If that were not enough, in 2014 CA allegedly started to experiment with the covert creation of pages and groups on Facebook and other social media testing campaign messages, without declaring the real owners and administrators, or their algorithmic work. When a critical mass of members or followers was reached, administrators and troll accounts would suggest physical meet-ups that would further invigorate, and often inflame targets, in a highly invasive process of behavioural change, without any real knowledge from the targets of the experiments they became part of “in the wild.”<sup>84</sup> We do not know whether such experiments were carried out during the 2016 Trump election campaign, but there is no reason to exclude such possibility.

Another aspect that Sumpter glosses over is that, even if the prediction rates of the models he tested were as low as his findings suggest – which might not necessarily be the case given that the data in CA possession was likely of a far bigger magnitude and perhaps better quality than the dataset for student training tested by Sumpter – in the end CA was not indiscriminately targeting all Americans. As mentioned before, according to CA internal report, they were targeting 9 million voters in several battleground states. With the distortions of the electoral college, they only needed to influence the voter behaviour of a relatively small fraction of people. Trump won because of 80,000 votes in Michigan, Pennsylvania and Wisconsin. That gave him enough delegates to win the election, even though he trailed behind Clinton by nearly 3 million votes in the national count.<sup>85</sup> Two of those three states (Michigan and Pennsylvania) were targeted by CA.

This means that, if the audience segmentation and big data analysis worked well enough, even a small increase in prediction rates over random chance could have shifted that amount of voters with enough money invested in

---

84 Wylie, *Mindf\*ck*, pp. 121–123.

85 This is because the US presidential election is won not by the candidate with the most votes in the ballot box, but by the candidate that wins at least 270 of the 538 electoral votes, which are distributed by states. The electoral votes assigned to each state are won as a block by the candidate winning the majority in that state, regardless of the margin of victory – this is true in all states except Nebraska and Maine which have a different way to allocate those votes. This system ends up with several distortions, such as the award of the same amount of electoral votes to states of different sizes, and the fact that battleground states where the outcome is most uncertain end up counting much more in the competition than states where the outcome is considered certain (e.g. traditionally Republican or Democratic states). So, as in the 2016 case, a small amount of votes in battleground states tends to determine the overall outcome. See also: Bump, P. (2016). Donald Trump Will Be President Thanks to 80,000 People in Three States. *Washington Post*, 1 December; Prokop, A. (2016). Why The Electoral College Is the Absolute Worst, Explained. *Vox*, 19 December.

internet ads. Beyond the figure provided by the CA internal report of around USD 5.6 mil, we know that the 2016 Trump campaign as a whole did spend big on Facebook targeted ads: USD 100 mil, according to the campaign digital director Brad Parscale.<sup>86</sup>

Interestingly enough, Sumpter himself makes the same point later in his book, citing approvingly another piece of research about voter turnout.<sup>87</sup> Bond and others ran a randomised controlled trial of political mobilisation messages on Facebook, in collaboration with the tech company.<sup>88</sup> Around 60 million US Facebook users were shown a message with information about their polling station and a button that users could press to indicate that they had voted. The message was accompanied by the faces of Facebook friends who had already pressed the button. The researchers selected a smaller group of users of around 600,000: these users were shown the same message, but without their friends' faces. There was also a control group of around 600,000 users who were shown no message at all. They tested the hypothesis that the "social message" (the one with the friends' faces) was more likely to bring people to the polls than the message without the social element of the friends' faces. They found that this difference was small (0.39%), but when scaled up through such big numbers thanks to the power of social media, the effective difference on voting behaviour was substantial: the researchers estimated that the social message might have brought 60,000 new voters to the polls through the direct effect of the messaging, and 280,000 more voters through the online social contagion effect, as the messages influenced not only the Facebook users directly targeted, but also their friends and friends of friends. In his analysis of these results, Sumpter concludes that "[a] small nudge made a big change to the number of people participating in democracy."<sup>89</sup> So why would the Cambridge Analytica case be any different, at least in principle? Why could CA not have influenced through their data campaigning a few dozens of thousands of votes that delivered Trump's victory in battleground states?

The 2012 study that Sumpter mentions in such a positive light reviews other studies of voter mobilisation, arguing for a position that fits the CA case rather well:

---

86 Loizos, *When You Spend*.

87 Sumpter, *Outnumbered*, pp. 147–148.

88 Bond et al. (2012). A 61-Million-Person Experiment in Social Influence and Political Mobilization. *Nature* 489: 295–298; Bond et al. (2012). Supplementary Information for "A 61-Million-Person Experiment in Social Influence and Political Mobilization," [http://fowler.ucsd.edu/massive\\_turnout\\_supplement.pdf](http://fowler.ucsd.edu/massive_turnout_supplement.pdf) (Accessed 29 October 2020).

89 Sumpter, *Outnumbered*, p. 148.

Voter mobilization experiments have shown that most methods of contacting potential voters have small effects (if any) on turnout rates, ranging from 1% to 10%. However, the ability to reach large populations online means that even small effects could yield behaviour changes for millions of people. Furthermore, as many elections are competitive, these changes could affect electoral outcomes.<sup>90</sup>

### Big Data, Human Insight and In-depth Qualitative Research

At another point in his book, Sumpter captures the essence of his argument: we are too easily swayed by the power of big data, and we often buy into hyped messages that emphasise the power of the machine, somewhat implying that it is the machine that provides insights and solutions, and not the humans working with it.

Commenting on the limitations of election polls and on the criticisms waged against popular election polling blog FiveThirtyEight's mistaken prediction of a Clinton victory in 2016, Sumpter concludes that:

Academic research has shown that polls are typically less accurate than prediction markets. As a result, FiveThirtyEight has to find a way of improving its predictions. There is no rigorous statistical methodology for making these improvements; they depend much more on the skill of the individual modeller in understanding what factors are likely to be important in the election. It is data alchemy: combining the statistics from the polls, with an intuition for what is going on in the campaign.<sup>91</sup>

The “data alchemy” Sumpter refers to is an important reminder that big data are used and manipulated by humans, and humans ultimately make the adjustments and the interpretations needed to make the models and the results useful and usable. But it would be incorrect to use this point as a criticism of the CA method.

Wylie's account confirms just that: it was CA employees, not the algorithms, that were using and abusing models and big data in conjunction with other methodologies, ultimately driven by their human insight. Wylie devotes several pages of his book reporting on in-depth qualitative data collection

90 Bond et al., A 61-Million-Person Experiment, p. 295.

91 Sumpter, *Outnumbered*, p. 99.

with Americans in focus groups but also unstructured everyday settings.<sup>92</sup> In a rather ethnographic fashion, Wylie ended up watching TV with some of the participants in their homes.<sup>93</sup> These activities allegedly informed the data-driven models that were going to be deployed to collect data about, predict and ultimately modify voter behaviour.

That is why the US fieldwork conducted in early 2014 was carried out by “sociologists and anthropologists, none of whom were American.”<sup>94</sup> Deploying a common argument held by many anthropologists, Wylie makes a paean to anthropology as a discipline that understands the deep otherness of societies such as the US, something that, according to the whistleblower, Americans themselves would not be able to do as uncritical insiders to their own society:

There’s a tendency among Americans to see their country as exceptional, but we wanted to study it like we would study any country, using the same language and sociological approaches. It was fascinating to explore America this way, and because I am not American myself, I felt I was more able to cut through unquestioned assumptions of American culture and notice things that Americans don’t see in themselves.<sup>95</sup>

Wylie goes on to exoticise Americans’ “fetishistic” attitude towards guns with a vignette from a trip to rural Virginia – in ways that are now seen as rather questionable within anthropology itself, but would feature well in older anthropological accounts.<sup>96</sup> A similar approach is used in other parts of the book to describe his encounters with right-wing Americans, as he and his colleagues were laying down the qualitative groundwork for big data modelling, personality profiling and other sophisticated forms of audience segmentation.

With its numerous references to specific concepts and findings from qualitative social science, Wylie’s narration is a clear example of the social theories and human insight behind big data. It makes sense, because by the time we start reading about the power of big data, we have been given the social scientific human-centred conceptual frameworks to see what CA scientists were looking for, and how they allegedly weaponised that large amorphous mass of data they had gotten hold of.

---

92 Wylie, *Mindf\*ck*, pp. 69–91.

93 *Ibid.*, pp. 75–76.

94 *Ibid.*, p. 87.

95 *Ibid.*, p. 87.

96 *Ibid.*, p. 88.



For anthropology and other cognate in-depth qualitative disciplines, the descriptions of CA journey through qualitative fieldwork cannot but sound alarming: so many ethnographers and other researchers carrying out in-depth qualitative studies will recognise elements of their own work, and yet, to see where our data and analysis could end up is no pretty sight. It provides another angle on dual use anthropology: not only the dual use of those like Wylie and colleagues who consciously went on to work for companies that weaponised traditional qualitative research methods, but also the unwitting dual use of the research of academics who are not working for such companies and interests, but who might one day see their scientific findings redeployed for highly unethical purposes.

Again, Wylie's story should be read in context. And the context here is provided by the qualitative methods SCL, CA parent company, had been using for decades. The basic insight behind Target Audience Analysis (TAA) might not be ground breaking, but perhaps this is exactly why the method seems feasible: if you want to change people's opinions and actions about something, do not bombard them indiscriminately with one message, but carry out in-depth studies of your target group, and then craft a variety of messages targeted at specific segments of that group.

Data is collected to find out the group's basic social, economic, psychological, political and cultural characteristics, and to understand its further articulation into subgroups with people holding different views and opinions, and various traits that can be exploited for behavioural change operations. As the 2016 Trump campaign and other populist campaigns across the world have shown us, you can target the most disparate groups with different messages that make them susceptible, in order to gather their support, or to suppress support for rival politicians.

The role of big data here is not that of some magical all-powerful tool that does away with human insight altogether, but rather of a powerful granular information infrastructure that can scale up and increase the effectiveness of such humanly crafted messages to reach out to people and social groups that would otherwise not be reached by conventional campaign methods. Two features make the US an ideal playground for this kind of operations: a very high internet and social media penetration that produces immense amounts of personal data of all kinds; and the absence of regulatory mechanisms protecting personal data capture by private companies, enabling companies to do as they wish with the large masses of personal data made available.

## Cambridge Analytica, Data Campaigning and the Future of the Human

To go back to the original question, my answer is that it is quite possible that CA could have effectively provided a significant advantage for Trump in his 2016 presidential election campaign. The discussion is relevant well beyond the confines of the 2016 US election and of the now defunct CA or its successors.

CA operations should not be seen as a sensationalised story of “rogue” scientists, but rather as an effort to advance common techniques of contemporary digital capitalism such as audience segmentation and microtargeting in the field of election campaigning, closely related to the relentless rendering of human experience into data for the dual purpose of commodifying and securitising potentially all aspects of humanity.<sup>97</sup>

If then companies like CA can carry out “persuasion” campaigns of such magnitude and depth, there is reason to worry: such methods are carried out largely with the lack of meaningful and true informed consent, and without affected and targeted actors knowing that they have been manipulated and enrolled for purposes that are not their own. Zuboff’s compelling treatment of “surveillance capitalism” (another label for digital capitalism) conceptualises this as the repurposing of human experience for others’ ends – in this case the others are the corporations, the colluded politicians and the technological cadres that actively support the infrastructure of digital capitalism.

Many questions remain unanswered and the debates I have drawn upon in this article are still in their nascent form. The epistemological, methodological and empirical insights of anthropology can be of great service.

We need to dig deeper into what it is that makes big data valid and applicable and under which conditions, and explore in more detail how humans are effectively reduced to data points and reconstructed into modelled versions of the “real thing.” Such an analysis would need to go back to key epistemological questions about the nature of modelling, about the status of qualitative vis-à-vis quantitative data, and about holism, complexity and discreteness. How are real people broken down into discrete units of data, and then recomposed in all kinds of assemblages through algorithms? What happens to humans and humanity in that process?

A subplot running through this article has to do with dual use and the role of anthropologists and other social scientists in furthering the scope and

---

97 Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books.

applicability of such data-driven tools. The question of ethical responsibility and the ongoing collusion between universities and military, intelligence and corporate interests is as cogent as ever.<sup>98</sup> We need to build on the crucial work carried out on these topics by anthropologists, but we also have to revise and adapt their insights to a different organisational and political economy context.<sup>99</sup>

The CA data scandal was of a quite different magnitude in terms of media coverage and popular reactions from the Wikileaks saga and Edward Snowden's NSA exposé. In the latter, governments have been active in prosecuting and categorising whistleblowers as criminals and traitors, while this did not happen with CA and SCL. If anything, the government actors who have been involved with CA and SCL refuse to claim responsibility for their actions, and the companies' status as private contractors creates a grey area of responsibility, where governments' doctrine of plausible deniability is pushed to new heights. The CA whistleblowers' accounts read more in the line of corporate professionals breaking ethical guidelines, than agents of military and intelligence interests interfering in domestic and foreign politics.

What is dual use when the boundaries between corporate and academic research are increasingly blurred, and industry applications are seen as less militarised than classic forms of recruitment from state military and security agencies? What kind of ethical smokes and mirrors, self-justifications and public perceptions are at play when somebody is hired by CA rather than the CIA?<sup>100</sup>

Finally, we should not lose focus on the more immediate goals of campaigns such as those carried out by CA for Trump in the US. This is probably the topic for another article, but the CA whistleblowers' accounts confirm that the CA case is as much about big data and dual use as it is about the weaponisation of racial, ethnic, gender, sexual and religious identities, with the return of "culture" to the centre-stage of political campaigning. Anthropologists have largely grown weary of the culture concept and of its essentialist connotations, but is it perhaps time to critically engage with "culture" again? Is Stroeken right when he claims that "the anthropologists' silence [on culture] condones for the larger public the hierarchy of cultures that is used to justify military intervention in

---

98 See also Laterza, Cambridge Analytica.

99 See for instance Price, *Cold War* on dual use in the cold war period, and González, *American Counterinsurgency* on the recent deployment of social scientists in "human terrain" studies.

100 On anthropologists working for the CIA during the Cold War, see Price, *Cold War*.

Iraq, Afghanistan, and soon Africa?”<sup>101</sup> Does this apply to today’s right-wing propaganda operations and their weaponisation of culture too? Why do so many ethnographies of contemporary far right and populist movements not focus on the media and communication infrastructure that produces right-wing populist propaganda?<sup>102</sup>

Without strict scrutiny and regulatory frameworks, the kind of methods used by CA are likely to be deployed in current and future political campaigns. Online persuasion operations have certainly not been put on hold because of the pandemic, but could in fact become tools to weaponise the virus for geopolitical and biopolitical agendas that go beyond the immediacy of the Covid-19 health threat. These tactics are also likely to thrive in a context of even deeper and wider internet penetration, as social distance and stay-at-home measures have pushed more and more people online and for longer parts of their daily life. In the 2020 US presidential election campaign, we have seen Trump’s public messages repeatedly mention Covid-19 as the “China virus,” further fuelling

---

101 Stroeken, K. ed. (2012). *War, Technology, Anthropology*. Berghahn, p. 2. Consider for instance political scientist Samuel P. Huntington’s notorious theory of the “clash of civilizations” as one example of a gross oversimplification of cultural difference that has been widely used to justify the military and security activities springing from the War On Terror launched by the US in the aftermath of the September 11 attacks; Huntington, S.P. (2011 [1996]). *The Clash of Civilizations and the Remaking of World Order*. Simon and Schuster; Haynes, J. (2019). *From Huntington to Trump: Thirty Years of the Clash of Civilizations*. Lexington Books.

102 On this point, see also the debate between myself and Chris Hann in Alexander, C. and Wardle, H. (2019). A Conversation Around Keith Hart: Swimming Into the Current of Human Society Through History. Member Voices, *Fieldsights*, 31 October, <https://culanth.org/fieldsights/swimming-into-the-current-the-movement-of-human-society-through-history> (Accessed 29 October 2020). One important case in this regard is that of Italy’s Five Star Movement (M5S), a populist political movement with strong nationalist and anti-immigrant overtones founded in 2009 by comedian and blogger Beppe Grillo and the late internet strategist Gianroberto Casaleggio. The movement started largely through online groups formed in the popular Meetup platform, and became the first Italian party in the 2018 general election. It has been in government since then. The internet company Casaleggio Associati (initially led by the late Gianroberto Casaleggio and now in the hands of his son, Davide) has been involved with the M5S since the start, and has been providing the technological infrastructure and digital communication expertise that launched the movement as a primarily internet-based movement, including developing the e-democracy platform Rousseau, where members vote on a range of issues and political decisions. See also: Galbo, J. (2020). Making Sense of Italian Populism: The Five Star Movement and the Lega. *Alternate Routes: A Journal of Critical Social Research* 31(1): 51–66; Laterza, V. (2018). Italy: The New Anti-Migrant National Consensus. *Daily Maverick*, 17 July.

his hostile geopolitics against the Asian superpower.<sup>103</sup> This was coupled with calls for “herd immunity” as a response to the pandemic – the proposal, rejected by most scientists, that letting the virus run through the community would build immunity and be a better alternative to the social and economic damages caused by lockdowns.<sup>104</sup>

This article and the recent work by Besteman and Gusterson, Boellstorff and Maurer, and Seaver, have hopefully shown the promise of an anthropological understanding of what Genevieve Bell calls “the socio-technical imagination” that underpins big data and algorithms.<sup>105</sup> With an emphasis on human agency and on the multiple human and non-human logics that span across technology and society, anthropologists are particularly well positioned to study the processes of continuity and change that sustain rapidly evolving human-technology relations. Anthropology can help us understand what it means to be human in today’s technological world and reflect on what the future has in store for humans as social, cultural, political and economic beings. Anthropologists should not miss the opportunity to contribute to these crucial debates. They should push the conversation further, within and beyond their disciplinary boundaries, and should engage with multiple audiences inside and outside academia.

### Acknowledgements

I am grateful for the generous insight of Louis Philippe Römer, Ivan Puga-Gonzalez, and Victor Adetula, who helped me think through some of the issues discussed in this article. Patience Mususa provided important feedback and comments on an advanced draft. Conversations with Lauren Paremoer and Divine Fuh helped shape the direction and content of the work. The article also benefited from the constructive comments of two anonymous reviewers.

<sup>103</sup> Remarks by President Trump in Press Briefing, September 16, 2020, <https://www.whitehouse.gov/briefings-statements/remarks-president-trump-press-briefing-september-16-2020/> (Accessed 29 October 2020).

<sup>104</sup> The Editorial Board (2020). COVID-19 Herd Immunity Strategy Fits Donald Trump’s Failures in Coronavirus War. *USA Today*; Greenhalgh, T., McKee, M. and Kelly-Irving, M. (2020). The Pursuit of Herd Immunity Is a Folly – So Who’s Funding this Bad Science? *The Guardian*, 18 October.

<sup>105</sup> Besteman, Gusterson, *Life by Algorithms*; Boellstorff, Maurer, *Data, Now*; Seaver, What Should an Anthropology; Bell, G. (2015). The Secret Life of Big Data. In: Boellstorff, Maurer, *Data, Now*, p. 9.