

5-2021

The Rhetoric of Big Data: Collecting, Interpreting, and Representing in the Age of Datafication

Brad Mehlenbacher
University of Waterloo

Ashley Rose Mehlenbacher
University of Waterloo



This work is licensed under a [Creative Commons Attribution-Noncommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/)

Recommended Citation

Mehlenbacher, Brad; and Mehlenbacher, Ashley Rose. "The Rhetoric of Big Data: Collecting, Interpreting, and Representing in the Age of Datafication." *Poroi* 16, Iss. 1 (2021): Article 3. <https://doi.org/10.13008/2151-2957.1311>

Hosted by [Iowa Research Online](https://www.iowa-research-online.org/)

This Article is brought to you for free and open access by Iowa Research Online. It has been accepted for inclusion in *Poroi* by an authorized administrator of Iowa Research Online. For more information, please contact lib-ir@uiowa.edu.

The Rhetoric of Big Data: Collecting, Interpreting, and Representing in the Age of Datafication

Acknowledgements

This article draws on research supported by the Social Sciences and Humanities Research Council Insight Grant program as well as the Ontario Early Researcher Award program. We would like to the editor of this special issue, Iulian Vamanu, and also Amanda Axley and the reviewers for their helpful feedback.

The Rhetoric of Big Data:

Collecting, Interpreting, and Representing in the Age of Datafication

Brad Mehlenbacher

*University of Waterloo
Waterloo, Ontario*

Ashley Rose Mehlenbacher

*University of Waterloo
Waterloo, Ontario*

Poroi 16,1 (May 2021)

Abstract: Rhetorical studies of science, technology, and medicine (RSTM) have provided critical understanding of how argument and argument norms within a field shape what is meant by “data.” Work has also examined how questions that shape data collection are asked, how data is interpreted, and even how data is shared. Understood as a form of argument, data reveals important insights into rhetorical situations, the motives of rhetorical actors, and the broader appeals that shape everything from the kinds of technologies built, to their inclusion in our daily lives, to the infrastructures of cities, the medical practices and policies concerning public health, etc. Big data merits continued attention from RSTM scholars as our understanding of its pervasive use and its ethos grows, but its arguments remain elusive (Salvo, 2012). To unpack the elusivity of big data, we explore one particularly illustrative case of big data and political, democratic influence: the Facebook-Cambridge Analytica scandal. To understand the case, we turn to social studies of data to explore the range of ethical issues raised by big data, and to examine the rhetorical strategies that entail big data.

Keywords: Big Data, Cambridge Analytica, Ethics, Expertise, Facebook, Rhetoric of Technology

Introduction

Big data describes a range of ideas and activities that are difficult to illustrate precisely, and the polysemy of the term opens a space for various analyses. In this paper, we describe some developments of the term, its practical use to achieve influence, and how a rhetorical vantage might reveal sometimes overlooked actors and agents



involved in the rhetorical world of “big data.” No doubt the business of big data is flourishing, despite its uncertain meaning across fields. In the midst of the COVID-19 pandemic, the question of data, its collection and use, and its meaning, as well as what promises big data might hold to help map the spread of infection are featured in daily briefings, the nightly news, and newspapers around the world. Data holds promise for which we currently so desperately hope and, surely, big data must provide still more of what we need. As we examine what data itself means, and how big data is distinct, we find a much more complex story than the quantitative connotations of the term promise. Data is well known to be biased as a product of systems that are biased, including by racism and sexism (Hicks, 2018; Noble, 2018, especially on big data beginning on page 29), and big data, too, replicates such problems (Eubanks, 2018). Data and big data do not simply reveal the world to us, but help to craft the world, and so those who construct those data are rhetorically significant actors. So, too, are those who are responsible for making data meaningful to problems. We might call such actors “data analytics experts” and their particular expertise is also an important key to understanding big data and its rhetorical significance and influence. In this essay, we explore how big data is rhetorically distinct from previous conceptions of data, and how the rhetorical actors and agents involved in big data may come to shape its meaning. Importantly, we also explore the rhetorics of big data itself and how its promises and perils entail old rhetorical and ethical problems. We do so by exploring what we hope is perhaps an outlier case, but a case that seems all too relevant for all its complexity and indeterminacy: Cambridge Analytica.

Cambridge Analytica was a U.K.-based consulting company that was involved with a large data breach of Facebook user data. The widely reported story of Cambridge Analytica follows that data from Facebook was used to help develop “profiles” of voters that could then be used to develop targeted political messaging. Messaging was shared through social media platforms and aimed at “persuadables” to encourage them to vote (or not) in a favorable direction to the clients of Cambridge Analytica (Amer & Noujaim, 2019). In then-Chief Executive Alexander Nix’s words, “... we were able to use data to identify that there was very large quantities of persuadable voters there that could be influenced to vote for the Trump campaign from this day forward” (Channel 4 News, 2018). At another level, however, big data as a concept itself seemingly became a tool of persuasion among clients while the actual practices of the company may have involved less than data-driven

approaches. For example, a Channel 4 News undercover story revealed the CEO of Cambridge Analytica suggesting the company had used tactics that may be considered entrapment (Channel 4 News, 2018). Cambridge Analytica's motto, "Data drives all that we do"¹ (Frankenfield, 2018, para. 2), it would seem, was something of an over-simplification and points to some of the concerning marketing and conversations surrounding big data. Later, after the Cambridge Analytica story broke and as the fallout generated discussions about data privacy, big data once again became a tool of persuasion, this time among publics, to suggest that the conversation ought to focus on who is collecting our data rather than, say, the persuasive tactics that may have been used in the U.S. and U.K. and before that with elections in Nigeria, Kenya, the Czech Republic, India, and Argentina (Channel 4 News, 2018).

Numerous aspects of this case deserve attention, but for the purposes of this essay, we wish to focus on the use of "big data" as a concept in Cambridge Analytica's work. Ultimately, lawmakers in both the U.S. and Britain called for further investigation of Cambridge Analytica's involvement in the U.S. 2016 elections and the Brexit referendum. The story is a complicated one, but an essential contention about the business model for Cambridge Analytica was their use of data collected from Facebook. As one engages more deeply in the complex case of Cambridge Analytica, which includes purported data breaches, concealing breaches, but also tactics as suggested in the Channel 4 undercover study, one begins to doubt the centrality of data analytics alone. In the case of Cambridge Analytica, locating data at the nexus of the scandal that in fact involves much older ethical problems in persuading audiences may divert attention from important questions rhetorical scholars can address. As Cambridge Analytica Global Political Managing Director Mark Turnbull points out in Channel 4 News' undercover video, "We just put information into the bloodstream of the Internet and then just watch it grow" (Channel 4 News, 2018).

Focus on the data and technological systems alone moves attention away from another narrative, one much less technologically impressive, but revealing of the broader milieu that shapes rhetorical situations where data-driven responses are enacted. Put simply, big data in this scenario *might* allow for a

¹ After Cambridge Analytica's motto, and for editorial consistency with the quotations we draw from in our case analysis, we use the word "data" as a singular noun throughout this article.

novel manner by which to classify voters and craft messaging, although this is debated (Szalai, 2019), but it is the socio-rhetorical situation that will inform how messages are crafted, not only big data insights. We wish in our analysis then to hold attention on two key facets of big data: its media form, and its influence on constructing a rhetorical situation and the attendant response. Before investigating one case that illustrates these two facets, some definitional work is necessary to reveal the underlying complexity of data and big data. We then discuss the role of big data ethics, ultimately focusing on the nature of data analytic expertise and the importance of truth and trust in experts working with big data.

The Meaning of Data and Big Data

Revealing the complexities of big data is usefully approached by first exploring the idea of data itself. Data, Rosenberg (2013a) reminds us, is not an immaterial but a material *thing*. Although it is common to think of data as a semiotic object of analysis, and indeed it might be, there is in fact a kind of materiality (in addition to its semiotic or representational comportment toward an “object in reality”) that makes data particularly persuasive. Understanding data as forming facts versus reflecting facts is less obvious than one might initially conclude. This is an issue that Rosenberg in his (2013a) chapter “Data before the Fact” illuminates using Mary Poovey’s discussion of facts versus data, when she writes “What are facts? Are they incontrovertible data that simply demonstrate what is true?” Rosenberg observes that, to Poovey, “Facts may be conceived either as theory-laden or as simple and incontrovertible” and, “In the latter case, we call them ‘data.’” (p. 17). However, this reasoning also opens a complicating line of inquiry such that, if facts are made up of multiple data, then data can potentially be viewed as “theory-laden” as well. Several related terms are worth pausing to consider here, notably “fact” and “evidence,” which Rosenberg (2013a) defines as follows:

The word “data” comes to English from Latin. It is the plural of the Latin word *datum*, which itself is the neuter past participle of the verb *dare*, to give. A “datum” in English, then, is something given in an argument, something taken for granted. This is in contrast to “fact,” which derives from the neuter past participle of the Latin verb *facere*, to do, whence we have the English word “fact,” for that which was done, occurred, or exists. The

etymology of “data” also contrasts with that of “evidence,” from the Latin verb *videre*, to see. (p. 18)

Taken together, Rosenberg explains that each of these terms provide important contributions to knowledge-making systems, arguing that “facts are ontological, evidence is epistemological, data is rhetorical” (Rosenberg, 2013a, p. 18). To understand how data is rhetorical, it is instructive to look back even earlier than the term’s Latin form. Data also has early specialist or technical meaning, which can be found in its Greek form (*dedomena*). The 4th century BCE mathematician Euclid used data as “givens” (what is known and can be used to solve problems), and this use of the term in mathematics would have considerable staying power. Throughout the early modern period, this meaning of data as a given remained in use, particularly in the fields of math and theology, Rosenberg explains, and in these fields, “*data* referred to things taken for granted and thus not inquired after” (emphasis original, 2013b, p. 560). In English, data would be imported through mathematics in the 18th century and would have retained its powerful persuasive appeal. Consistent across Latin and vernacular uses in English, *data* did not refer to an objective, out-there, capitalized truth, but rather to “claims accepted *for the sake of argument*” (Rosenberg, emphasis original, 2013b, p. 561). Rosenberg explains, “What made such claims ‘data’ was their rhetorical status, not some intrinsic formal or material quality: in another argument, the very same statements might not be ‘data’ at all” (2013b, p. 561). The data was, this is to say, situated by the particular argument or rhetoric within which an argument unfolded. But as data was adapted from a Latin word into the vernacular English, it transformed in an important manner. Data began to be used in the singular in English, a shift from its plural Latin form (Rosenberg, 2013a, p. 19), and with this shift the “English language provides tools for discussing ‘data’ as *something*” (Rosenberg, emphasis original, 2013b, p. 562). Further, the ideas concerning “what constituted *givenness* were changing,” coming to “evoke a particular sort of representational entity upon which one could operate through systems of calculation, classification, and communication, while holding the question of referential truth in abeyance” (Rosenberg, emphasis original, 2013b, pp. 565–566). Articulating data in this way provides us with a framework for meaningfully talking about big data, which Emmanuel and Stanier (2016) describe as “the collection, processing, analysis and visualization associated with very large data sets,” and as “an umbrella term to cover a range of data, technologies and applications” (p. 1). Big data then is not just data

per se but instead operates as a complex system of objects, methods of collection and analysis, and tools for visualizing. Whereas Rosenberg (2013a) observes that we “want to give data an essence, to define what exact kind of fact data is,” in our present moment “it may be that the data we collect and transmit has no relation to truth or reality whatsoever beyond the reality that data helps us to contrast” (p. 37).

From Rosenberg’s historical analysis of the word “data” we find several developments essential to understanding contemporary uses of the term. Most notably, we observe the manner by which data once operated rhetorically as situated with respect to a particular argument, and how the word transformed from the plural Latin form through vernacular use in English as either plural or singular, which in turn affords the possibility of a materiality to the givenness of data. Before turning to the materiality of data, we first want to explore the matter of givenness and the rhetoricity of data. The importance of understanding the situatedness of data in its ancient or modern sense, with respect to argument, is what makes data an object of rhetorical inquiry. After Boyd and Crawford (2012), we acknowledge that big data is more than a technological fact; rather, big data comprises a cultural, technological, and scholarly phenomenon that highlights technology (“computational power and algorithmic accuracy”), analysis (enabling pattern identification “to make economic, social, technical, and legal claims”), and mythology (“promise of higher form of intelligence and knowledge ... with aura of truth, objectivity, and accuracy”) (p. 663). And the interplay of these phenomena is part of the problem of big data: if we cannot guess at the power of big data in terms of tracking and capturing people’s behaviors, we are susceptible to unsubstantiated claims about the power and objectivity of numbers to predict and influence human behaviors. As Iliadis and Russo (2016) warn, “Big Data must remain open to cultural, ethical, and critical perspectives, particularly when viewed as a modern archive of data facts and data fictions. Data, along with its sciences and infrastructures, are informed by specific histories, ideologies, and philosophies that tend to remain hidden....” (p. 2).

Definitions you might find of big data—were you to glance at an industry website—might not highlight these features but do include both technical and non-technical features. Beginning with some examples of the technical features, SAS (n.d.) writes “Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis,”

and Oracle Canada (n.d.) explains “Big data is data that contains greater variety arriving in increasing volumes and with ever-higher velocity. ... big data is larger, more complex data sets, especially from new data sources. These data sets are so voluminous that traditional data processing software just can’t manage them.” IBM (n.d.) defines big data as “the use of advanced analytic techniques against very large, diverse data sets that include structured, semi-structured and unstructured data, from different sources, and in different sizes from terabytes to zettabytes,” adding also some of the characteristics of big data, including “high volume, high velocity or high variety.” Big data is structured, unstructured, complex, new. SAS (n.d.) also defines big data by its purpose: “But it’s not the amount of data that’s important. It is what organizations do with the data that matters. Big data can be analyzed for insights that lead to better decisions and strategic business moves.” And Oracle Canada (n.d.) notes, “But these massive volumes of data can be used to address business problems you wouldn’t have been able to tackle before,” while IBM (n.d.) tells us that “Analysis of big data allows analysts, researchers and business users to make better and faster decisions using data that was previously inaccessible or unusable.” Approaches are used, according to IBM (n.d.), “to gain new insights from previously untapped data sources independently or together with existing enterprise data.” Thus, we learn that there are multiple purposes to which big data might be put, for example in the financial industry (Tett, 2018), regulation (Groenfeldt, 2020), and in tracking COVID-19 symptoms (Yasinski, 2020). Big data can simultaneously be viewed comparatively in terms of relational and other systems views such as graph theory or object orientation, or big data can be viewed in terms of attributes—that is, the three Vs, volume, velocity, and variety (Gandomi & Haider, 2015; Uprichard, 2013)²—or in terms of its architectural or environmental constraints, where we think about big data in terms of horizontal scaling and parallel processing and, importantly, as being non-relational in nature (Emmanuel & Stanier, 2016, pp. 2–3). Data then is both analytically generated and analyzed by data analysis experts. And thus, after Rosenberg (2013a), we argue that

² As a construct, the idea of big data is often discussed in terms of the “3Vs” of volume, velocity, and variety, and other lists of V-words (Gandomi & Haider, 2015; Uprichard, 2013) and even P-words such as political, partial, and predictive (Lupton, 2015), but Kitchin (2013, 2014) and Kitchin and McArdle (2016) note the fundamental importance of qualitative distinctions.

“it is this rhetorical aspect of the term ‘data’ that has made it indispensable” (p. 37).

In asking “what makes big data *big*” we find ourselves frequently being asked to interrogate both the quantitative and qualitative qualities that make big data distinct from regular data. When Kitchin and McArdle (2016) ask “What makes Big Data, Big Data” in their article, they provide an answer that demands rhetorical inquiry. First, they chart a similar argument we have made here: big data, as an idea, is nebulous (cf. Favaretto, De Clercq, Schneble, & Elger, 2020). Definitions often focus on the size of a dataset, the conditions of its construction and creation, and perhaps its structure, but not the ontological characteristics of big data. Kitchin and McArdle provide an important intervention identifying this erasure of big data’s quiddity as a central theoretical question. Through a survey of previous studies in big data, they identify several characteristics that are notably distinct in what we might call big data from smaller datasets. Comparability is an interesting characteristic that distinguishes big data, providing perhaps “greater comparability between countries,” but also has potentially greater challenges in terms of representativeness, reporting errors, missing data, and outliers (Kitchin & McArdle, 2016, p. 3, Table 2). Biases, too, are a matter of distinction: we do not know them. Sampling errors and non-sampling errors are notable, as well; Kitchin and McArdle write, “missing data, reporting errors and outliers although possibly less frequently occurring, and new types of errors” (Kitchin & McArdle, 2016, p. 3, Table 2). A significant volume of data exists relative to other sources; these data might be national or international, and the demographics might be more difficult to measure. Intellectual property is yet another consideration. *Purpose*, however, is the most notable of the distinguishing characteristics from a rhetorical vantage. Survey data, for instance, is “Designed for statistical purposes” and administrative data is “Designed to deliver/monitor a service or program” but big data is “Organic (not designed) or designed for other purposes” (Kitchin & McArdle, 2016, p. 3, Table 2). *Purpose* is a central concern in distinguishing big data from other kinds of data. *Purposelessness* of some form of big data, too, can be a distinguishing factor. When Kitchin and McArdle identify purpose as one distinction in big data, writing that its purpose is “organic (not designed) or designed for other purposes,” the concept of big data is interestingly complicated (2016, p. 3). If we cannot identify the purpose for which data was identified, isolated, and then collected, that is, the rhetorical conception of data as a kind of

argument, then new tools for understanding the rhetorical work accomplished by big data are required.

Another way to think about purpose might be to ask about the motive behind the data collection, as well as the analysis and use, and these questions entail another theme: expertise. As Hartelius (2018) writes of big data in her rhetorical analysis of the United Nations Global Pulse program mission to integrate big data into humanitarian development efforts, “Of primary concern regarding big data are, in tandem, expertise and motive” (p. 68). Protagoras’s *homo mensura* doctrine, Hartelius argues, is illustrative of how data and big data are ontologically determinate through the perceptual gaze of the human. Hartelius writes that “digital data about humans in the world exists only when, and because, it is collected *as data*” (emphasis original, p. 70). The United Nations Global Pulse program, Hartelius explains, uses big data to set policy agendas, and for that reason it is critical to understand the function of data in that “Information becomes the foundation for deliberation and judgement, both of which can serve good or evil” (2018, p. 68). The sophistic understanding of why we measure and what that means for how we create knowledge in the world also moves us away from conflating data with positivistic notions of fact or truth to be discovered in the world. Hartelius (2018) explains,

Protagoras’s perspective on measurement and understanding is strikingly relevant: Humans are compelled to measure, to pursue knowledge. Acting on that compulsion, we not so much discover an external reality as we invent a structure, or a technology, that reflects our own positionality and agenda. That is, we find what we were looking for; not because it was ‘out there’ but because that is what we knew to seek. (p. 69)

Big data in some ways gives us more to look through, and perhaps better tools for looking through, but it remains true that knowing what to seek is a key to making the enterprise of “analytics” a useful one. Floridi (2012), too, explains the importance of seeking, noting that big data is not well-defined and alluding “to an overwhelming sense that we have bitten off more than we can chew” (p. 436). He explains that it is not the overabundance of data that is a problem since it is a “resource” (p. 436); rather, it is important to know what one is looking for in that data. Floridi echoes Plato’s *Cratylus*, writing that “the game will be won by those who ‘know how to ask and answer questions’” (qtd. in Floridi, 2012, p. 437).

Appropriately, Floridi explains who will do this asking and answering: “such epistemological skills are taught and applied by a black art called analytics” (2012, p. 437). In this art, the job is to seek and not find, but invent. By invent, we invoke the rhetorical sense of invention, to oppose the idea of merely discovering what is “out there,” as Hartelius observes (p. 68). Because invention is critical to this work, looking to those who have the capacity for such invention is central to understanding the purpose of big data and the ethical entailments of such inventional work. Giardullo (2016) sees this as the pragmatic work of all social science researchers, to carefully choose among alternative methodological choices in the design of the research where methodological questions about big data are essentially methodological questions about data in general (p. 544). Having unpacked the complexity of defining big data and before we examine the nature of expertise and ethics in such analytic work, we turn to an illustrative case, Cambridge Analytica.

Cambridge Analytica and Purpose

It was an international scandal involving numerous governments and nations’ elections that brought to our attention again to the ways in which social media data—as big data—might be used to change our behaviors beyond the products we buy. The Cambridge Analytica scandal revealed, and continues to reveal, the pervasive way some of our most fundamental public and private institutions and our democratic systems are threatened by persuasive appeals delivered to our feeds without our knowledge, consent, or transparency. What began as discussions about sophisticated tools for analysis and exclamations about the power of big data turned gradually to warnings about purposeful misinformation and descriptions of the dangers of information fragmentation. Most of the individuals involved in the scandal strongly deny any wrongdoing, even as scientific researchers and the general public grapple with the realization that their personal data can be employed in ways they might not be comfortable with and that might reveal personal information they did not intend to share (Anonymous, 2018).

The London-based company Cambridge Analytica reportedly used 87 million users’ data (Kang & Frenkel, 2018) that “was obtained and used without permission” from Facebook to build voter profiles (Rosenberg, Confessore, & Cadwalladr, 2018). Indeed, as far back as 2014, Cambridge Analytica gave John Bolton—Trump’s eventual national security advisor—and his super PAC a prototype of its

Facebook data-driven profiling technology. And what the company promised in return was to build a set of tools that could “identify the personalities of American voters and influence their behavior” (Rosenberg, Confessore, & Cadwalladr, 2018). CBC News described how Cambridge Analytica claims to know a great deal about online users: the company can “combine standard data-mining techniques (e.g., geography, age, gender) with personality modeling, so called psychographics (e.g., education, shopping, interests, etc.)” (Larsen, 2018). In addition to data mining, another layer of user information is reportedly added, one that is “about personality, decision making, and motivation” (example words they then referred to included “openness, conscientiousness, extraversion, agreeableness, and neuroticism”) (Larsen, 2018).

One early source of insider information about how Cambridge Analytica was data mining, personality modeling, and narrative targeting was then-28-year-old whistleblower Christopher Wylie, a Canadian data consultant who worked for the company from 2013 to 2014. Wylie, who has described himself as “one of the creators of Cambridge Analytica” (Wylie, 2019), “a founder” (Blatchford, 2018), and as “director of research” (Cadwalladr, 2018), argued that the company was a data-mining “arsenal of weapons” in a culture war (Wylie, 2019). Sharing his experiences working for Cambridge Analytica when he testified before the U.S. Congress in June, 2018, Wylie began his story when he joined Strategic Communication Laboratories (SCL), a company that he claimed was “supplying the U.K. Ministry of Defense and NATO armies with expertise in information operations” (Wylie, 2019). Specifically, the company gained access, according to Schneble, Elger, and Shaw (2018), to “320,000 user profiles and their friends’ data through the ‘thisisyourdigitallife’ app” developed by Dr. Aleksandr Kogan, an assistant professor of psychology and neuroscience at Cambridge University, U.K. Bartlett (2018) and Patrikarakos (2019) estimate that Cambridge Analytica had between 2,000 and 5,000 data points on 240 million Americans over eighteen years of age.

After the scandal broke in March, 2018, with the publication of Cadwalladr and Graham-Harrison’s (2018) story in *The Guardian* and Rosenberg, Confessore, and Cadwalladr’s (2018) story in *The New York Times*, an interesting follow-up to the story involved a professor, Dr. David Carroll, at New York’s Parsons School of Design who employed a British data protection law to sue Cambridge Analytica to find out what data the company collected about him. Notably, it was not Carroll’s idea to sue Cambridge

Analytica; it was first the idea of Paul-Olivier Dehaye, a researcher in Geneva, Switzerland, who runs a nonprofit named Personaldata.IO. Dehaye was investigating SCL's claims about the influence of its data-collection methods on the Brexit referendum outcome in the U.K. (Lapowsky, 2019), and he even paid the 10-pound fee for the request for data disclosure that David Carroll submitted. If Carroll was unable to win the appeal to view the data that Cambridge Analytica had collected on him, he reasoned that Americans would be able to see how citizens of the U.K. had certain legal rights to privacy under U.K. law. Initially, Carroll received an Excel file from the data compliance team at Strategic Communication Laboratories (Cambridge Analytica's parent company) that contained information on where Carroll lived, how he voted, and how much he cares on a scale from 0 to 10 about issues like the national debt and gun rights. Carroll pushed for the specific data they had collected about him, and Cambridge Analytica refused to provide it. Ultimately Carroll's lawyers filed a court order for Cambridge Analytica and their parent company, SCL (O'Sullivan, 2018). Although he won the case, Carroll never received data from SCL or Cambridge Analytica, even as the companies filed for insolvency proceedings in May 2018 and are now defunct. Channel 4 News (2020) did recently obtain what they referred to as "the Trump Campaign database" and David Carroll's data file, a file that contained his "personality scores"—openness, conscientiousness, extroversion, agreeableness, and neuroticism scores—in addition to "Lifestyle and Spending Habits" presented as variables such as "Avid Gamers Model = 8," "Impulse Buyer Model = 9," or "Cat product model = 3." Carroll concluded, "This is what we were fighting to get; this is what I knew was there" (Carroll qtd. in Channel 4 News, 2020); however, the data does not tell us what models were used to generate numerical values for each variable and it is still not clear how the personality scores informed message microtargeting or if they even did.

There are numerous layers of concerning ethical behavior operating here, most of which are still being investigated and discussed.³ It is perhaps useful to begin early in the development of these profiles. First, Kogan collaborated with two colleagues,

³ As of 5 October 2020, though, the U.K.'s Insolvency Service did disqualify Alexander Nix, ex-CEO of Cambridge Analytica, from "holding directorships, or from promoting, forming or managing a company" for seven years due to his "potentially unethical" role in the big data scandal (Davies, 2020).

Michal Kosinski and David Stillwell of the Psychometrics Centre at Cambridge University on the data-mining algorithms used in the psychological testing app; Kosinski and Stillwell were reportedly excluded from the collaboration with SCL, the parent company of Cambridge Analytica and in 2015 when Kogan applied for university ethics approval to use the data he collected, his application was denied (Lewis, Grierson, & Weaver, 2018). This is not entirely surprising because the app in question was designed to look like a personality test but, when downloaded, obtained access to the user's profile information and that of their friends (Szalai, 2019). Another issue was that users were able to give consent for access to their friends' data as well as their own. Facebook's terms of service do not allow sharing of data with third-party vendors; however, allowing app designers to collect user data is not against its terms of service (Hertwer, 2018). Facebook "demanded that the company simply delete the data" but they did not admit to the public that they knew about the data violation or about the company's actions (Schneble, Elger, & Shaw, 2018). Ultimately, there would be more concerns about the engagements of Cambridge Analytica than their use of Facebook data, reminding us that the story is of more than the power of big data. Wylie has been quoted as saying it "frustrates me ... how dominant the Facebook angle of the story was, when there's so much [obscenities]ed-up [obscenities] that Cambridge Analytica were doing in different parts of the world" (Cadwalladr, 2019). Adding frustratingly more complexity to the case, Peter Pomerantsev reports in his (2019) book, *This is Not Propaganda*, that Nigel Oakes, the founder of SCL, told Pomerantsev that "he was skeptical that Facebook 'likes' and online product purchases could replicate months of in-depth, in-field research" (p. 184). Oakes's original model for the branch of SCL would be rebranded as Cambridge Analytica's (p. 184).

Virtues and Vices in Big Data Ethics

In traditional data ethics there is considerable attention given to the humans impacted by the collection and use of data. In the case of Cambridge Analytica, much public debate on this scandal surrounds issues of privacy, consent, and research ethics concerning human subjects. Some questions are related specifically to who consented to the use of Facebook data, for use by which actors in the scandal, and if participants knew the ends to which their data would be put. We also acknowledge the importance of privacy and resisting surveillance. For example, there are several concerns, such as surveillance or exacerbation of inequalities,

raised with large-scale COVID-19 mobile tracking apps in the U.S. used for contact tracing (Frith & Saker, 2020). And the science and technology of contact tracing apps are less than settled. As British Columbia's provincial health officer Dr. Bonnie Henry has noted, when considering contact tracing apps, it is important to "find the right IT support for the work that we're doing that doesn't create more problems than it solves" (Canadian Press, 2020). That is, given that the practice of contact tracing is already well-established in scientific expert communities, technical solutions need to augment those processes rather than providing, as Henry states, "generic messages about who they may or might have been in contact with" (Canadian Press, 2020). Another important question concerning big data is: how do we ensure ethics approval or, at least, ethical use for repurposed data? In the case of Cambridge Analytica, much of the data set that their analysts used to generate user "universes" was legally obtained, through purchases of credit-card behaviors, consumer spending data, phone surveys, and so on (Bartlett, 2018). As well, Kogan, the Cambridge University professor who co-constructed the personality test app that harvested Facebook data without obtaining permission, was also unable to obtain permission from Cambridge University's ethics board to collect the data in the first place. We cannot fully understand the rhetorical possibilities of big data nor the kinds of experts charged with creating meaning from those data without understanding the broader ecology within which data is generated.

There are indeed actors we might consider beyond the human in this broader ecology. Human-generated content can be used to inject divisive cultural materials to encourage partisan polarization, voter apathy, and perhaps worse. As the content is circulated, people "like" or share the content, and data about these circulations and engagements are once again integrated back into the big data set. Big data is, in this way, recursive. Big data is produced by user-generated inputs, but big data produces the conditions for those user-generated inputs. Behind the scenes there are certainly human agents put "information into the bloodstream of the Internet" as quoted above from the undercover Channel 4 documentary (2018). Yet, to borrow a metaphor of our pandemic time of writing, what has been injected can be understood as an infectious agent, which now spreads among hosts, reliant upon the human subject but changing that subject, spreading beyond that subject, and evolving in a broader techno-human ecology of social and online media platforms. Big data, in this way, acts as a *thing* somewhat outside of the control of human actors. Put another way, the data is borne of a

different rhetorical situation; it takes shape not merely as constructed arguments within a particular context in big data situations, but rather comes to the situation as an actor, shaped by previous circumstances, and acts both by its original (now decontextualized) design and also as nodes within a network that can respond unpredictably. Mauthner (2019) argues for a post-humanist position on big data ethics by eschewing positivist perspectives, discovery-based epistemologies with the hopes of revealing truths to be found in the data, or even humanist notions of socially constructed reality.

Following the work of Barad (2007), Haraway (1988), and others, Mauthner argues that the posthuman ethical framework “shifts the focus away from *the power of researchers over research participants* toward the “*world-making*” powers of practices of inquiry: their ability to constitute the very nature of their objects/subjects of study” (2019, p. 671). What is considered to be “unethical” in this framework is “claims to innocently represent the world ‘as it really is’” (Mauthner, 2019, p. 671). Such an approach is an important addition to thinking on data ethics insofar as it extends the range of matters to be considered. Still, we disagree with some of the characterizations in this account that posit that a humanist ethics is one that “takes the rational human subject as both the locus of moral agency and object of ethical concern” (Mauthner, 2019, p. 674). Indeed, the rhetorical tradition has long challenged the idea of the *rational* human subject and attendant problems with that language, and is more comfortable with the charge of “intentional human subjects” (if not conflated with rational as Mauthner does) as a focal point of rhetorical-humanist inquiry (Mauthner, 2019, p. 679). Mauthner’s approach offers a lens to examine the way that (big) data comes to shape our world through practices in knowledge-making, and this is a key consideration when we ask what happens once the results of analytic work are “injected” online. Mauthner explains,

The object of ethical concern is not humans but the world-making powers of practices of inquiry—their ability to bring specific configurations of the world into existence. It is in this sense that ethics is seen as inseparable from knowledge making and is understood as responsibility and accountability for the performative effects of knowledge practices—where responsibility and accountability are conceptualized in other than humanist terms. (2019, p. 681)

We find this framing helpful as a way to consider a variety of problems that this case presents beyond the most obvious (and, we add, quite serious) matters of humanistic concern, including privacy and autonomy. Specifically, we are interested in those questions of the “world-making powers” that might be circulating online. How do platforms participate in rhetorical appeals (Kelly [now Mehlenbacher], Kittle Autry, & Mehlenbacher, 2014; Moriarty & Mehlenbacher, 2019), and how do those objects—or, *things*—act upon us and with us to in turn continue knowledge-making? We are also interested in questions that take an environmental perspective concerned with the impacts of energy demands that big data and supporting computational systems require. Particularly, we are interested in inquiring after the ethical consequences of unnecessary cycles and annual terawatt hours being committed to big data computing storage and analysis, especially if big data cannot deliver the insights being promised. If we were to answer these questions we might also ask where responsibility lies (Johnson & Johnson, 2018). To answer these questions, however, is not possible with the current status of this case, or in fact our own expertise. Indeed, the case itself is not transparent, and the science is not settled, so instead we pose these questions as broader themes to address in the study of big data. Such questions are important, and so too is taking a less human-centric approach to understand those ethical entailments. But we here return to humans because it is those rhetorical actors who have obfuscated the case in such a manner that it becomes a challenge to move into these posthuman questions. Because we lack access due to the proprietary nature of the data and systems, and lack technical expertise along with expertise in the psychological sciences, we, like many other non-experts in these domains, look to those brokers of specialist or technical knowledge to understand the actors, the stakes, the situation, and the appeals: the experts.

Indeed, to understand moral dimensions of big data, it is critically important to examine key actors as they work to rhetorically frame the discourse. Certainly, as the Cambridge Analytica case has shown, accessing the data itself or the technological mediations that entail the data collection and analysis process is not always possible. Indeed, inaccessibility due to claims to propriety likely make this a significant challenge in the study of big data. Even where we might gain access, there is another challenge for users, namely that many are not knowledgeable or expert in these systems. Understanding how one’s data is used requires sophisticated technical understanding, and even among purported

experts, there is debate. There are also numerous questions about who has the appropriate kind of specialist or expert knowledge in terms of technical solutions and application to domain-specific problems. For Cambridge Analytica, it is not merely the technical proficiency in collecting user data that is a key part of the story. Additionally, part of the story is the psychological profiles that were developed (and their veracity) as well as the attendant but distinct claims about what those profiles afford in terms of persuasive capacities. In such cases there is always a danger of ultracrepidarianism, that is, of analysis by people who are not experts on the big data or user profiling, and it is not merely the general expert to whom we ought to look for answers, but the expert in the specific problem area we wish to address.

Data Analytic Experts

Appeals to big data do not only rely on the data, but on the idea of experts behind the scenes turning information into something *meaningful*. The appeal is one of an expert *ethos*, where some specialized knowledge or skills are used to perform what seems indistinguishable from magic: build a profile based on an array of data not only to really gain insight into the minds of individuals, but also to determine how to configure those individuals into groupings that one might then craft targeted messaging to in an effort to increase sales or, in the Cambridge Analytica case, to persuade voters. The appeal to being able to perform such feats is one that relies on expert status (*ethos* is, of course, distinct from expertise). But do those who claim such expert status have the goods, the skills to perform such tasks, and if they do, what are the requisite skills? Being able to collaborate with other experts on big data projects is one important ability (González-Bailón, 2013). Technical skills to design or operate databases and systems (e.g., Hadoop) are certainly part of the repertoire that some of those involved in big data projects will need, but we might think of most people operating with these skills at the level of mastery but perhaps not expertise. Beyond technical skills, there is a matter of data analytics, or making meaning of the data. Floridi (2012) identifies among these skills the ability to identify “small patterns” (p. 436). As a feature of big data, it is not the traditional model of inventive potential that the *topoi* of data structures, databases, ontologies, etc.,⁴ provide but rather a different sort of *idioti topoi*—

⁴ Invention with *topoi* has been identified as a key approach to the construction of scientific (e.g., Walsh [now Olman], 2010) and technical

they are neither common topics nor disciplinary knowledge per se, but rather something between. Specialized and situated, but not domain-specific, necessarily.

Seeming purposelessness in big data (recalling Kitchin & McArdle, 2016) then is a matter of rhetorical vantage and acumen. For experts in this realm, the apprehension of meaningful small patterns is generative of purpose. Purpose is not an operation in discovery but rather a matter of invention. And that is precisely the kind of work involved in creating an app that simulated a psychology test but in actuality obtained permission to harvest the user's Facebook profile data and the data of their friends' accounts. Analytics experts at Cambridge Analytica were then able to build models of users and then design and user-test ads specially designed for these "universes" of user types, or "psycho-graphics" (Bartlett, 2018). Further, crafting personas or understanding the composition of one's audience and their personalities or character is not itself new.

University of Cambridge research news published an article in 2011 entitled "Cambridge researchers have created a website that combines the Facebook profiles of fans of companies and public figures with personality testing to create what they are describing as a 'revolutionary' new marketing tool." Kosinski and Stillwell, Kogan's two colleagues, remember, conducted work that is foundational to but distinct from Kogan's dealings with Cambridge Analytica. The locus of expertise, the expert *ethos*, and the expert insights that Cambridge Analytica claims to offer are illuminated by Kosinski and Stillwell's research. The article discusses a new website, LikeAudience.com, which "was created by Michal Kosinski and David Stillwell, both researchers at the University of Cambridge's Psychometrics Centre" (University of Cambridge, 2011, para. 7). Explaining how different Facebook page likes can offer insights into personalities, Kosinski writes,

For the first time, it means that companies, politicians, celebrities and anyone else with a Facebook presence can

knowledge (e.g., Roundtree, 2016). For example, Hartzog (2015) examines the *Aedes aegypti* mosquito in VectorBase and its Dengue Fever Ontology (IDODEN) and argues that the organization and standardization of the database operates as a "tool of rhetorical invention" by crafting topoi (following Walsh's [now Olman's], 2010, model) from which arguments might be drawn from the ontologically (computational) "linked" or structural data (p. 10).

investigate not just how many people “like” them—they can also draw up a detailed profile that includes information about their average follower’s personality, IQ and satisfaction with life. Other data such as the gender balance and average age of their fanbase is also made available. (University of Cambridge, 2011, para. 4)

The data, Kosinski states, comes from an app they created that provided a personality test along with these Facebook data. Using the “Big Five” personality traits (i.e., conscientiousness, extraversion, openness, agreeableness, and stability, with continuums within these), the app generates data combined with “[o]ther apps, such as an IQ test ... All of this data is then combined with the participant’s Facebook profile information—such as their age, hometown, and relationship status” (University of Cambridge, 2011, para. 12). The data is the sell:

LikeAudience’s creators believe that it will be of particular value to marketers, who will be able to uncover new potential audiences for their advertising campaigns, and exploitable niches based on the fans of their closest rivals. The potential significance for politicians, particularly when on the election trail, is also clear—although it can throw up some interesting results. (University of Cambridge, 2011, para. 14)

Let us consider two of these profiles offered as examples in the article from Cambridge University research news, which appear in an appendix-like section to the article with the title “10 profiles from LikeAudience”:

2. Sarah Palin

Who likes her? People with a strong sense of tradition, and in particular people who like order, structure and self-discipline. Palin appeals to older people with high life satisfaction and, for a self-confessed “hockey mom”, her Facebook fanbase has a male bias.

What do they like? Former US President George W. Bush, the patriotic “United States of America Fan Page”, Pizza Hut and the Seattle Seahawks. (University of Cambridge, 2011 paras. 21–23)

And, as a second, perhaps less political and more market-based example,

9. Radiohead

Who likes them? Average fans exhibit high openness, which suggests they are imaginative, curious, creative and sensitive to beauty. They also have high IQs and not many friends compared with your typical Facebook user. Coincidentally, this fits the profile of a paranoid android.

What do they like? The film adaptation of Hunter S. Thompson's *Fear And Loathing In Last Vegas*; William Golding's novel *Lord Of The Flies*, Tarantino's *Kill Bill* series, and *Monty Python And The Holy Grail*. (University of Cambridge, 2011, paras. 42–44)

We chose these examples for their range, but it is certainly worth examining all 10 in an effort to understand the patterns in these profiles. What we can see here, however, is that the profiles are not entirely surprising. Even the seemingly insightful appeal that Palin's fanbase skews male is not surprising: A CNN survey in 2008 reported that men had a more favorable opinion of Palin (Steinhauser, 2008). As a profile, the analysis of Radiohead is particularly interesting in the obscure reference to "fit[ting] the profile of a paranoid android." Several references mark certain associations that might also be unsurprising if one were to look at the data from a generational perspective. "Kill Bill" and "Fear and Loathing in Las Vegas" were released around the same time and Radiohead was quite popular at the time, so there is likely a group of people that came of age during that period who this would have appealed to. And this demographic would have likely read *Lord of the Flies* in high school and "Monty Python and the Holy Grail," although predating Radiohead, likely fit as a part of the cultural package for certain teenagers or young adults at the time.

Kogan and business partner Joseph Chancellor, recall, were said to have excluded Kosinski and Stillwell from the negotiations with SCL (Lewis, Grierson, & Weaver, 2018). Later, Kogan would tell Sumpter (2018) that he did not believe the correlation between likes and personality were particularly strong, and specifically cites CEO Nix as not understanding what Kogan's work did, saying "Nix has very little comprehension of what he is talking about," and that "He is trying to promote [the personality algorithm] because he has

a strong financial incentive to tell a story about how Cambridge Analytica have a secret weapon” (Kogan qtd. in Sumpter, 2018, para. 15). Of Wylie, the so-called whistleblower, Kogan told Sumpter “He is speaking outside of his expertise. He’s not a data scientist. At SCL, he dealt with business development and data law. He had no role I know of in handling data and certainly no role in modeling” (Kogan qtd. in Sumpter, 2018, para. 21). Some of those outside the debate, looking in, are not convinced about the science behind this approach either, and some are not even convinced about the quality of predictions that data from Facebook itself can help us make (Broad, 2018). Gibney (2018) outlines in *Nature* some of the scientific debate in psychology and allied areas of research about the effectiveness of micro-targeting based on personality in marketing and voter influence in comparison to other forms of targeting. The findings are less impressive than the claims made by Cambridge Analytica (at times) and the numerous stories reporting the company’s data activities. We say “at times” because CEO Nix himself has said, when brought before U.K. Parliament, that Kogan’s research on Facebook data in the creation of personality profiles “proved to be fruitless” (Nix qtd. in Gibney, 2018, para. 16) and, when speaking with NPR, that “hundreds of thousands of Americans” were surveyed as the basis of Cambridge Analytica’s profiles (Nix qtd. in Gibney, 2018, para. 17). Marketing researchers took to *The Conversation* to share their field’s qualifications on the grandiose claims of big data’s power to craft personality profiles that will have much success, explaining that research dating back to the 1990s illustrated that “abstracted and universalised personality type cannot capture the complexity and cultural sensitivity of consumer lifestyle choices, symbolic expression and tastes” and that “This shift in thinking put an end to the wider application of psychographic methods long ago, at least in the field of marketing and consumer research” (Rokka & Airoidi, 2018, paras. 11–12). Kosinski, though, disagrees and still sees dangers in the approach. Sumpter spoke with Kosinski and observes “[h]e sees the correlation between Facebook likes and psychologists’ personality scores as a side-effect of the deeper understanding algorithms have of us than we have of ourselves” (Sumpter, 2018, para. 25).

What is the legitimacy of these claims to expertise? We don’t know. It is, however, compelling to learn that researchers and specialists in psychology, marketing research, and high tech itself (Fish, 2018) all question the validity of the claims concerning how data was used and the likelihood that such approaches themselves

will be successful to the degree that Cambridge Analytica seemingly banked on. Still, their persuasiveness to clients relied on their purportedly data-driven approach—their “secret weapon”—and consistently drove Cambridge Analytica’s public persona through the company motto and name (cf. Confessore & Hakim, 2017). The appeal to expertise in this case is concerning because the idea of the expert is one that is necessarily relational, meaning that part of being an expert is being recognized as such by colleagues working in similar or related areas. The Cambridge Analytica case thus contributes to an undermining of trust among (purported) technical experts and publics. Although we might dismiss matters here as a singular case, the broader concerns about what technologies are capable of and the ethical concerns raised in this essay are much more pervasive than would allow such an easy casting aside. Illustrated in this case is the importance of understanding how broad expert appeals on the grounds of technical specialism can work to redirect attention from the speaker’s *ethos* to the presumed *logos* of the data. However, as we have argued, data, big data, and the entailments surrounding big data, are not merely matters for reason, or fact, etc., but rather firmly in the domain of rhetorical appeals. Little consensus among experts in the particular case of Cambridge Analytica points to a wider debate about big data and its applications—indeed, its meaning, too.

Such uncertainty is a problem when the stakes are as high as functioning democracies. Indeed, scientists are held to a higher standard, including among rhetoricians. As Ceccarelli (2020) recently argued, extending her work with Pietrucci (Pietrucci & Ceccarelli, 2019), scientists “have an ethical obligation to communicate about risks they are uniquely qualified to recognize, and to do so in a way that makes their knowledge accessible to their fellow citizens of the world who would otherwise be forced to act without all they need to make good decisions” (p. 241). Here we have a case where not only have the experts (as such) failed to do so, but, further, created conditions to undermine the ability of citizens to make good decisions. As Wilson (2020) argues in an analysis of Facebook and the importance of *ethos* and dwelling, “it is the connections we have in our ethe ecologies that we trust, not technologies that were created for purposes against our best interests” (p. 227). Among those connections, broadly, we might include expert voices.

Final Remarks

Data and other technology are only part of the story in this case study—as are the elements of big data and broader technological tools—and a rhetorical account of the human elements are, too, critical. As Delfanti and Frey (2020) note in their study of Amazon’s efforts to automate their warehouses, using humans to extend machines, “[h]uman labor—the input it provides to machinery and the values it generates—must remain at the center of analyses of automation,” and we would say that, explicitly, of big data, too (p. 21). Pausing to look at the rhetorical appeals that are ultimately made—after all the promises of impressive, big data, sophisticated computer models and profiles, psychographics, user universes, and micro-targeting—Cambridge Analytica Global Political Managing Director Mark Turnbull’s words reveal the ancient understanding that drives the messaging. Turnbull’s argument, captured in the Channel 4 News undercover footage, is that the motivations of the average voter are “the two fundamental human drivers, ... hopes and fears, and many of those are unspoken and even unconscious,” adding:

You didn’t even know that was a fear until you saw something that just evoked that reaction from you and our job is to drop the bucket further down the well than anybody else to understand what are those really deep-seated underlying fears, concerns. There is no good fighting an election campaign on the facts because actually it’s all about emotion. (Turnbull qtd. in Channel 4 News, 2018)

Aristotle told us that fear (*phobos*) can be defined as “a sort of pain or agitation derived from the imagination” (*phantasia*), and that fear is future-oriented in that it anticipates “destructive or painful evil” (Arist. *Rhet.* II.5.1, 1382a, trans. Kennedy). Further, Aristotle noted the importance of fear in deliberative speech, telling us that “fear makes people inclined to deliberation; while no one deliberates about hopeless things” (Arist. *Rhet.* II.5.13, 1383a, trans. Kennedy). Echoing across the millennia, these words are comported for concerning ends when applied to swaying voters today, but their rhetorical insight is not lessened. Indeed, the rhetorical effectiveness of such appeals, when unrestricted by norms and values held to be central to democratic engagement (such as truth, honesty, mutual understanding, trust, etc.), may be all the more so effective. Preserving our personal privacy and our

data rights then must also come with a call to preserve our rhetorical tools for assessing the claims about what these technological processes can do as well as the degree to which they can and do deliver. Are we receiving data-driven, technologically-designed, micro-targeted messages that are the product of a technological system's deep insights into ourselves, or are the messages rhetorical appeals we already know and effective not because of the hardware we have built and the data we have amassed, but more so because of the wetware within which we each reside and our relational, social, and, indeed, moral comportment toward one another? As we consider this case, we turn to Aristotle, who himself offered something akin to "profiles," wherein he spoke of the character of the young, the old, those in their prime, the powerful, and so on. In Book II of *Rhetoric*, of those affected by *dynamis* (power), Aristotle, while describing their virtues, cautions of their vices: "if they commit wrong, they do it on a large, not a small, scale" (Arist. *Rhet.* II.17.4, 1391a, trans. Kennedy).

Acknowledgments

This article draws on research supported by the Social Sciences and Humanities Research Council Insight Grant program as well as the Ontario Early Researcher Award program. We would like to thank the editor of this special issue, Iulian Vamanu, and also Amanda Axley and David Depew, and the reviewers for their helpful feedback.

Copyright © 2021 Brad Mehlenbacher and Ashley Rose Mehlenbacher.

Reference List

- Amer, K., & Noujaim, J. (2019). *The Great Hack* [Film]. Netflix.
- Anonymous. (2018). Digital trust: A scandal over an academic's use of Facebook data highlights the need for research scrutiny. *Nature*, 555, 29 March. 559–560.
- Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.
- Barasch, A. (2018, March 21). The Cambridge Analytica whistleblower isn't a hero. He's just another tech bro. *Slate*. <https://slate.com/technology/2018/03/the-cambridge-analytica-whistleblower-is-just-another-tech-bro.html>
- Bartlett, J. (2018, March 20). Big data is watching you: The Cambridge Analytica row shows politics moving in a disturbing direction. *Spectator*. <https://www.spectator.co.uk/article/the-cambridge-analytica-row-shows-politics-moving-in-a-disturbing-direction>
- Blatchford, A. (2018, March 20). Facebook whistleblower pushed data-mining boundaries in Canada: source. *CBC*. <https://www.cbc.ca/news/politics/wylie-cambridge-analytica-liberals-1.4583810>
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication, & Society*, 15(5), 662–679.
- Broad, E. (2018, March 22). Are Cambridge Analytica's insights even that insightful? *The Guardian*. <https://www.theguardian.com/commentisfree/2018/mar/22/are-cambridge-analyticas-insights-even-that-insightful>
- Cadwalladr, C. (2018, March 18). 'I made Steve Bannon's psychological warfare tool': Meet the data war whistleblower. *The Guardian*. <https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump>

Cadwalladr, C. (2019, March 17). Cambridge Analytica a year on: “a lesson in institutional failure.” *The Guardian*.
<https://www.theguardian.com/uk-news/2019/mar/17/cambridge-analytica-year-on-lesson-in-institutional-failure-christopher-wylie>

Calwalladr, C., & Graham-Harrison, E. (2018, March 17). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*.
<https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>

Canadian Press. (2020, May 6). B.C. health officer weighs in on contact-tracing apps. *The Globe and Mail*.
<https://www.theglobeandmail.com/canada/british-columbia/video-bc-health-officer-weighs-in-on-contact-tracing-apps/>

Ceccarelli, L. (2020). The polysemic facepalm: Fauci as rhetorically savvy scientist citizen. *Philosophy & Rhetoric*, 53(3), 239–245.

Channel 4 News. (2018, March 19). Data, Democracy and Dirty Tricks. *Channel4*. <https://www.channel4.com/news/data-democracy-and-dirty-tricks-cambridge-analytica-uncovered-investigation-expose>

Channel 4 News. (2020, September 29). ‘It works as a suppression system, it works to subvert the will of the people’—Professor David Carroll. *Channel4*.
<https://www.channel4.com/news/it-works-as-a-suppression-system-it-works-to-subvert-the-will-of-the-people-professor-david-carroll>

*Confessore, N. (2018, April 4). Cambridge Analytica and Facebook: The scandal and the fallout so far. *The New York Times*.
<https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>

Confessore, N., & Hakim, D. (2017, March 6). Data firm says ‘secret sauce’ aided Trump; Many scoff. *The New York Times*.
<https://www.nytimes.com/2017/03/06/us/politics/cambridge-analytica.html>

- Davies, R. (2020, September 24). Former Cambridge Analytica chief receives seven-year directorship ban. *The Guardian*. <https://www.theguardian.com/uk-news/2020/sep/24/cambridge-analytica-directorship-ban-alexander-nix>
- Delfanti, A., & Frey, B. (2020). Humanly extended automation or the future of work seen through Amazon patents. *Science, Technology, & Human Values*, 1–28. <https://doi.org/10.1177/0162243920943665>
- Emmanuel, I., & Stanier, C. (2016). Defining big data. *BDAW'16: Proceedings of the International Conference on Big Data and Advanced Wireless Technologies*. ACM.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. Picador.
- Favaretto, M., De Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PLoS ONE*, 15(2). <http://doi.org/10.1371/journal.pone.0228987>
- Fish, G. (2018, March 25). Computer scientist: Cambridge Analytica are dirty tricksters with an overhyped data operation. *Medium*. <https://medium.com/rantt/computer-scientist-cambridge-analytica-are-dirty-tricksters-with-an-overhyped-data-operation-7cb4ad290a31>
- Floridi, L. (2012). Big data and their epistemological challenge. *Philosophy & Technology*, 25(4), 435–437.
- Frankenfield, J. (2018, March 29). Cambridge Analytica. *Investopedia*. <https://www.investopedia.com/terms/c/cambridge-analytica.asp>
- Frith, J., & Saker, M. (2020). It is all about location: Smartphones and tracking the spread of COVID-19. *Social Media + Society*, July-September, 1–4.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144.

- Giardullo, P. (2016). Does 'bigger' mean 'better'? Pitfalls and shortcuts associated with big data for social research. *Quality and Quantity*, 50(2), 529–547.
- Gibney, E. (2018, March 29). The scant science behind Cambridge Analytica's controversial marketing techniques. *Nature*. <https://www.nature.com/articles/d41586-018-03880-4>
- González-Bailón, S. (2013). Social science in the era of big data. *Policy & Internet*, 5(2), 147–160.
- Groenfeldt, T. (2020, July 22). Clouds bring new risks to financial services. *Forbes*. <https://www.forbes.com/sites/tomgroenfeldt/2020/07/22/clouds-bring-new-risks-to-financial-services/#76e9f4131ab1>
- Haraway, D. (1988). Situated knowledges. *Feminist Studies*, 14, 575–599.
- Hartelius, E. J. (2018). A Protagorean analysis of the United Nations' Global Pulse. In M. Kennerly and D. S. Pfister (Eds.), *Ancient rhetorics and digital networks* (pp. 67–87). The University of Alabama Press.
- Hartzog, M. (2015). Species boundaries in genetic engineering: The case of *Aedes Aegypti* and dengue fever. *Bridging divides: Spaces of scholarship and practice in environmental communication*. COCE.
- Hertwer, N. K. (2018, May 1). Tech giants grapple with user privacy and misinformation. *Information Today*. <http://newsbreaks.infotoday.com/NewsBreaks/Tech-Giants-Grapple-With-User-Privacy-and-Misinformation-124536.asp>
- Hicks, M. (2018). *Programmed inequality: How Britain discarded women technologists and lost its edge in computing*. MIT Press.
- IBM. (n.d.). *Big data analytics: Leverage the most effective big data technology to analyze the growing volume, velocity and variety of data for the greatest insights*. <https://www.ibm.com/analytics/hadoop/big-data-analytics>

- Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, 3(2), 2053951716674238.
- Johnson, M. A., & Johnson, N. R. (2018). Can objects be moral agents? Posthuman praxis in public transportation. In K. R. Moore & D. P. Richards (Eds.), *Posthuman praxis in technical communication* (pp. 121–140). Routledge.
- Kang, C., & Frenkel, S. (2018, April 4). Facebook says Cambridge Analytica harvested data of up to 87 million users. *The New York Times*.
<https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html>
- Kelly [now Mehlenbacher], A. R., Kittle Autry, M., & Mehlenbacher, B. (2014). Considering Chronos and Kairos in digital media rhetorics. In G. Verhulsdonck & M. Limbu (Eds.), *Digital rhetoric and global literacies: Communication modes and digital practices in the networked world* (pp. 226–246). IGI Global.
- Kennedy, G. A. (2007). *Aristotle on rhetoric: A theory of civic discourse: Translated with introduction, notes and appendices*. Oxford University Press.
- Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in human geography*, 3(3), 262–267.
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.
- Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3(1), 1–10.
- Lapowsky, I. (2019, January 25). One man's obsessive fight to reclaim his Cambridge Analytica data. *Wired*.
<https://www.wired.com/story/one-mans-obsessive-fight-to-reclaim-his-cambridge-analytica-data/>
- Larsen, K. (2018, March 20). Who is Christopher Wylie? How a B.C. high school dropout set out on path to political data

harvesting. *CBC*. <https://www.cbc.ca/news/canada/british-columbia/christopher-wylie-background-1.4583054>

Lewis, P., Grierson, J., & Weaver, M. (2018, March 24). Cambridge Analytica academic's work upset university colleagues. *The Guardian*.
<https://www.theguardian.com/education/2018/mar/24/cambridge-analytica-academics-work-upset-university-colleagues>

Lupton, D. (2015). The thirteen Ps of big data. *This Sociological Life*.
<https://simplysociology.wordpress.com/2015/05/11/the-thirteen-ps-of-big-data/>

Mauthner, N. S. (2019). Towards a posthumanist ethics of qualitative research in a big data era. *American Behavioral Scientist*, 63(6), 669–698.

Moriarty, D., & Mehlenbacher, A. R. (2019). The coaxing architecture of Reddit's r/science: Adopting ethos-assessment heuristics to evaluate science experts on the Internet. *Social Epistemology*, 33(6), 514–524.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Oracle Canada. (n.d.). *What is Big Data?*
<https://www.oracle.com/ca-en/big-data/what-is-big-data.html>

O'Sullivan, D. (2018, March 18). New York professor sues Cambridge Analytica to find out what it knows about him. *CNN*. <https://www.cnn.com/2018/03/17/politics/professor-lawsuit-cambridge-analytica/index.html>

Patrikarakos, D. (2019, October 11). Book World: In “Targeted,” data is the precious coin of the realm of digital robber barons. *The Washington Post*.
https://www.washingtonpost.com/outlook/in-targeted-data-is-the-precious-coin-of-the-realm-of-digital-robber-barons/2019/10/11/f1cff630-eaa0-11e9-9306-47cb0324fd44_story.html

- Paytoncular, C. E. (2019, December 19). *Cambridge Analytica main players: Where are they now?* Women in Technology. <https://www.women-in-technology.com/wintec-blog/cambridge-analytica-main-players>
- Pietrucci, P., & Ceccarelli, L. (2019). Scientist citizens: Rhetoric and responsibility in L'Aquila. *Rhetoric and Public Affairs*, 22(1), 95–128.
- Pomerantsev, P. (2019). *This is not propaganda: Adventures in the war against reality*. Hachette Book Group.
- Rokka, J., & Airoidi, M. (2018, April 2). *Cambridge Analytica's 'secret' psychographic tool is a ghost from the past*. The Conversation. <https://theconversation.com/cambridge-analyticas-secret-psychographic-tool-is-a-ghost-from-the-past-94143>
- Rosenberg, D. (2013a). Data before the fact. In L. Gitelman (Ed.), *"Raw data" is an oxymoron* (pp. 15–40). The MIT Press.
- Rosenberg, D. (2013b). Data as word. *Historical Studies in the Natural Sciences*, 48(5), 557–567.
- Rosenberg, M., Confessore, N., & Cadwalladr, C. (2018, March 17). How Trump consultants exploited the Facebook data of millions. *The New York Times*. <https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>
- Roundtree, A. K. (2016). Startup weekends: Invention as process for proto-entrepreneurs. *2016 IEEE International Professional Communication Conference (IPCC)* (pp. 1–14). IEEE.
- Salvo, M. J. (2012). Visual rhetoric and big data: Design of future communication. *Communication Design Quarterly*, 1(1), 37–40.
- SAS. (n.d.). *Big Data: What it is and why it matters*. https://www.sas.com/en_ca/insights/big-data/what-is-big-data.html

- Schneble, C. O., Elger, B. S., & Shaw, D. (2018). The Cambridge Analytica affair and Internet-mediated research. *EMBO Reports*, 19(8), e46579.
- Steinhauser, P. (2008, September 9). Men's support gives Palin edge in latest poll. *CNN*.
<https://www.cnn.com/2008/POLITICS/09/09/palin.poll/>
- Sumpter, D. (2018, April 22). My interview with Aleksandr Kogan: what Cambridge Analytica were trying to do and why their algortihm [sic] doesn't work. *Medium*.
<https://medium.com/@Soccermatics/my-interview-with-aleksander-kogan-what-cambridge-analytica-were-trying-to-do-and-why-their-f869ef65d945>
- Szalai, J. (2019, October). Tales from inside a political data harvester. *The New York Times*, C5(L).
- Tett, G. (2018, July 12). The Cambridge Analytica scandal echoes the financial crisis. *Financial Times*.
<https://www.ft.com/content/b21ffb20-85b1-11e8-96dd-fa565ec55929>
- University of Cambridge. (2011, April 22). *With friends like these...* Internet Archive.
<https://web.archive.org/web/20130904202230/http://www.cam.ac.uk:80/research/news/with-friends-like-these...>
- Uprichard, E. (2013). Focus: Big data, little questions? *Discover Society* (1), 1–6.
<https://discoversociety.org/2013/10/01/focus-big-data-little-questions/>
- Walsh [now Olman], L. (2010). The common topoi of STEM discourse: An apologia and methodological proposal, with pilot survey. *Written Communication*, 27(1), 120–156.
- Wilson, N. (2020). Algorithmic dwelling: Ethos as deformation in online spaces. *Rhetoric Review*, 39(2), 216–229.
- Wylie, C. (2019). *Mindf*ck: Inside Cambridge Analytica's plot to break the world*. Verbena Limited.

Yasinski, E. (2020, July 21). Big data and collaboration seek to fight Covid-19. *The Scientist*. <https://www.the-scientist.com/news-opinion/big-data-and-collaboration-seek-to-fight-covid-19-67759>