The *Reality* of Data Mining: Sculpting Discourse, Knowledge, and the New Subject

Quinn MacNeil

Inception

This paper was written in the winter of 2019 for Dr. Matthew Flisfeder's "Critical Studies of Social Media" class.

Abstract

Debates over privacy are especially common in the digital age. They often materialize into attitudes of indifference with moralist cores purporting the question's irrelevance to the good, law-abiding citizen. While there are plenty of arguments to oppose this position on surveillance, this paper focuses on a different concern in the privacy debate—data mining. In this paper, I argue that data mining—that is the collection of information on the individual such as preferences, locations, emotions, interests, behaviour, demographic, etc.—has concrete effects on our realities. It does so by curating what is sensible and intelligible through discourse, through proxies that culturally embed "truths," and by constructing new subjectivities. Contrary to the position articulated above, I argue we should care deeply about privacy over our data and scrutinize the normalization of its being collected as a by-product of our participation on web 2.0, smart devices, and an ever-growing digital life.

Introduction

Data mining has become the cash-cow of the Web 2.0 landscape. This monetization has severely altered the structure of how social media businesses are run, as chief computer scientist Peter Eckersley explains, "if you're using Facebook, you're entrusting the company with records of everything you do… [and their] business model is to amass as much first-party and third-party data on you as possible, and slowly dole out access to it" (Singer). The transcendence of time and space has been one of the primary attractions of the internet and is especially tied to the features provided by Web 2.0—making it a revolutionary tool in human communication. Web 2.0, in this way, is improving our lives through things that were previously seen as impossible, but it is also changing the actions that get carried out in the world, the things that are perceived as true, and therefore, is changing our realities. As Taina Bucher wrote in her book *If…Then: Algorithmic Power and Politics,* "platforms act as performative intermediaries that participate in shaping the worlds they only purport to represent." (Introduction). To understood how non-human interference is shaping our world, Web 2.0 is a necessary object of critical analysis to explore who/what has the power to set the conditions for what can be known, and with what possible effects. This paper will examine three ways that reality is being shaped through non-human intervention: the curation of what is sensible and intelligible through discourse, the ability of proxies to culturally embed "truths," and the construction of new subjectivities. Throughout, I use the data mining and algorithmic processing practices of Facebook and social networking sites as a sustained case study of the point.

Data and Discourse

In *Radical Technologies: The Design of Everyday Life,* Adam Greenfield defines data as "facts about the world, and the people, places, things and phenomena that together comprise it, that we

collect in order that they may be acted upon" (210). The actions produced by data collection establish an element of power because actions shape real-world circumstances. A notable example of real-world effects based on data mining is the Cambridge Analytica scandal, an event where 50 million Facebook users' data was used to target voters on behalf of the Trump campaign during the 2016 presidential election (Roose 2018). The data was collected through an app called "thisisyourdigitallife," built by academic Aleksandr Kogan, where thousands of users were paid to take a personality test and agreed to have their data collected for academic use. However, the personal data of the Facebook friends of every individual who downloaded the app was additionally collected. These users were then targeted with information to dissuade them from voting, or information that preyed on their prejudices and anxieties to nudge them toward acting in a way beneficial to the Trump campaign. This power over the actions of citizens in the material world, substantiated by data mining, is at the heart of concerns of people like Carole Cadwalladr who, when discussing the role of the tech company AggregateIQ in Brexit, expressed: "This is Britain in 2017. A Britain that increasingly looks like a 'managed' democracy. Paid for by a US billionaire. Delivered by Facebook. And enabled by us."

To explain how algorithmically presented information plays a role in shaping reality, I will apply Michel Foucault's Discourse Theory to Facebook. Foucault's idea of discourse is: "a group of statements which provide a language for talking about a particular topic at a particular historical moment" (Hall 44). It is through discourse that objects of knowledge are produced, as nothing meaningful exists outside of discourse. This is not to say that nothing physically or conceptually exists outside of discourse, but rather, for something to be a perceptible object of knowledge, it requires a description through discourse. The creation of objects of knowledge helps structure reality by providing the parameters for the acts that can be performed. Taina Bucher, Professor of Communication and IT at the

University of Copenhagen, identifies that the classifying, sorting, predicting, and processing of data makes algorithms political[1] in that they present different information that strengthens some realities, while weakening others, resulting in certain actions/inactions being carried out in real life. Facebook, in this case, is the medium in which knowledge about an individual is generated. Every instance of participation on the platform provides the platform with new data on the individual: their likes/dislikes, eating habits, frequented locations, and most importantly, how to continuously captivate their interest and induce their participation further. The data being collected on the individual creates knowledge that shapes their social field by changing how the institution with the "knowledge" views them, which in turn, shapes the reality that gets presented to the them (an exercise of power). In this way, power shapes reality through its connection to knowledge and its ability to produce certain forms of acting and knowing, as outlined in the case of Cambridge Analytica. Power and knowledge mutually structure each other. As Bucher explains, the power and politics of algorithms condense/construct the conditions for what is "intelligible and sensible." And as mentioned before, Foucault outlined that what is intelligible and sensible, or what "makes sense," is how our realities and truths are formed. In summary, data is directly connected to power because data is categorized to produce patterns/meaning which create knowledge, which then open up new relations of power, which are immanent to the field of action.

Proxies and "Knowledge"

Proxies, or stand-in data, are another way in which our conceptions of truth, and therefore our actions in the world, change reality. In his book, *We are Data: Algorithms and the Making of our Digital Selves,*

[1] Bucher explains that her use of the word "political" is not in reference to parliamentary politics, elections, campaigns, or political communication in the strictest sense, but rather, as "the practices and capacities entailed in ordering and arranging different ways of being in the world" (3).

John Cheney-Lippold explains the way that metadata is used to kill people with drone strikes (39). In 2008 the US government loosened its wartime drone guidelines which "reindexed *terrorist* into *'terrorist'*: a pre-identified 'signature' of behaviour that the U.S links to military activity" (Cheney-Lippold 40).

Categorization/profiling is problematic as Mireille Hildebrandt explains in the chapter "Defining Profiling: A New Type of Knowledge." Hildebrandt defines profiling as "a set of technologies, which share at least one common characteristic: the use of algorithms or other techniques to create, discover or construct knowledge from huge sets of data" (17). She explains that the technologies (hardware), and techniques (software), "are integrated into profiling practices that allow both the construction and application of profiles" (18). This idea is echoed by Cheney-Lippold and the "as if" way of measuring a "terrorist" "who still looks and sounds very similar to whom the U.S. government has historically declared to be a terrorist (resident of the Middle East or neighboring nations, speaks Arabic or Urdu, non-white, and practices Islam)" (41). In this way, human prejudices become encoded into technology and additional patterns found in conjunction with the prejudices become incorporated into the algorithm as "truth" or the "recipe" for the sought-after object (O'Neil). The same is true for proxies which stand-in for missing abstractions within the algorithm, which then become a characteristic of the desired object despite it having no ontological truth. As Ed Finn explained in "What is an Algorithm?", "computationalism argues that algorithms have no ontological claim to truly describing the world but are highly effective at solving particular technical problems" (22).

Proxies help shape reality when algorithms use data that isn't rooted in truth, but rather data that has come to be understood as wisdom and has been embedded into our culture as such. Hildebrandt emphasizes that the "knowledge" gained from algorithmic profiling is problematic when it is applied to new categories of applications and

services—a kind of dogma used to make additional decisions. This idea is echoed by John Cheney-Lippold who challenges the level of truth in models that use abstract ideas (which inevitably have variables/stand-in information), which he refers to as a "construction," which then serve to construct other abstracts, thereby perpetuating false characteristics (45). By basing what comes to be the "object of belief" on flawed or false representations of the world, we are then re-sculpting the world through actions informed by such beliefs. Hildebrandt identifies profiling as an inductive way of generating knowledge and argues that as such, it should be perceived more as hypotheses than truth. This form of "knowledge" also calls into question the logic behind inductive reasoning which Enlightenment philosopher, David Hume, challenged because they are justifications for methods that predict "instances of which we have had no experience [to] resemble those of which we have had experience" (Henderson).

New Subjectivities

In *Digital Technologies of the Self,* Yasmine Abbas and Fred Dervin argue that through the collection of data, smart devices produce new differences and identities (25). They explain that through the "memorization of 'shallow' daily practices, 'only deeds, not thoughts', novel domains for self analysis and 'a whole field of experience…which was earlier absent' (Foucault 1988, 28)" opens up (25). They provide an example of smart devices that monitor daily practices such as how someone likes their coffee in the morning and what temperature they enjoy for their showers, and where a family may never have known they had slightly different preferences for these practices, they now have very specific profiles for these day-to-day matters (25). When these smart devices are able to communicate with one another, they gain predictive capabilities and begin to compose identities based on the patterns found among people who have similar preferences and predicts the unknown preferences of one individual based on the preferences it does

know. By producing these new differences and identities that were previously unavailable and an impossible avenue of knowledge production, this data "constitutes consumer-subjects with very specific preferences, desires, and appetites" (26). They conclude that the subject-shaping capabilities of smart technology can be a source of concern:

> …when the shower starts anticipating your preferred temperature based on the kind of coffee you just had. Of course, the temperature might happen to be precisely the one that you like, but when you lose track of the lines along which you are manipulated into a certain identity, tinkering with the constitutive facticity of your identity becomes difficult. (Abbas & Dervin 26)

Bucher draws on the philosopher Ian Hacking to explain the idea of "making up people" through the datafication Abbas and Dervin articulated above. Bucher writes, "'society became statistical through the "enumeration of people and their habits" and these new categories shape the way subjects interact with each other (10). In *The Psychic Life of Power,* Judith Butler identifies the subject as not only something that opposes power, but something dependent on power for its very existence: "if, following Foucault, we understand power as *forming* the subject as well as providing the very conditions for its existence and the trajectory of its desire, then power is [also]… what we depend on for our existence and what we harbour and preserve in the beings that we are" (2). Algorithms and data mining, in this way, create the subjectiviation of the entrepreneurial self, which in turn, changes our actions in the world and thus the reality we are a part of. As Alice Marwick explains, the infiltration of market-logic into everyday social relations, or neoliberalism, is modelled by Web 2.0 through creating neoliberal selves and rewarding those who adopt such subjectivities (6). The neoliberal self is constituted by new measurements and categorizations which were before unknown and is sustained because the "self production of the neoliberal subject is pleasurable for the subject" (Marwick 201). The neoliberal self is a code of conduct, as well as a product of the new

power relations opened up through Web 2.0, that informs the way we perceive and act in the online, as well as offline, world.

Conclusion

Data mining and algorithmic processing are changing our realities. Through setting the parameters for what is intelligible and sensible and the actions available to us through these parameters, the "knowledge" produced and reproduced through proxies which embed falsehoods into our collective notions of truths, to the creation of new subjectivities which open up new ways of experiencing the world; our reality is being shaped. Data mining provides the conditions for what gets accepted as an object of knowledge, incorporated as pseudo-knowledge in the case of proxies, and the knowledge that constitutes and informs the self, and for these reasons, we must be highly critical of the knowledge we produce. With these new bodies of knowledge, new power relations open up, and such forms of power—transnational, networked, secretive, and digital, tend to elude public scrutiny. I am not making the case that the reality-shaping capabilities derived from data mining and algorithms are inherently evil—after all, with every new technology, our actions in the world change. However, when new data-driven power relations open up, the channeling of cultural production in particular directions deserves proper scrutiny. It is imperative to ask why one form of power is being privileged, and if there is an alignment between those who hold power and the interests that are being served.

Work Cited

Abbas, Yasmine and Fred Dervin. *Digital Technologies of the Self*. Cambridge Scholars Publishing, 2009.

Bucher, Taina. *If ... Then: Algorithmic Power and Politics*. Oxford University Press, 2018.

Butler, Judith. *The Psychic Life of Power: Theories in Subjection*. Stanford University Press, 1997.

Cadwalladr, Carole. “The Great British Brexit Robbery: How Our Democracy Was Hijacked.” *The Guardian*, 7 May 2017, www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy.

Cheney-Lippold, John. *We Are Data : Algorithms and the Making of Our Digital Selves*. New York University Press, 2017.

Finn, Ed. *What Algorithms Want: Imagination in the Age of Computing*. MIT Press, 2017.

Foucault, Michel, et al. *Technologies of the Self: A Seminar with Michel Foucault*. University of Massachusetts Press, 1988.

Greenfield, Adam. *Radical Technologies: The Design of Everyday Life*. Verso, 2017.

Hall, Stuart. *Representation: Cultural Representations and Signifying Practices*. Sage Publications, in association with the Open University, 1997.

Henderson, Leah, "The Problem of Induction", The Stanford Encyclopedia of Philosophy (Spring 2019 Edition), Edward N. Zalta (ed.), plato.stanford.edu/archives/spr2019/entries/induction-problem/.

Hildebrandt, Mireille. “Profiling: From Data to Knowledge.” Datenschutz Und Datensicherheit - Dud, vol. 30, no. 9, 2006, pp. 548–552, *SpringerLink* doi:10.1007/s11623-006-0140-3.

Marwick, Alice Emily. *Status Update: Celebrity, Publicity, and Branding in the Social Media Age*. Yale University Press, 2013.

O’Neil, Cathy. “Weapons of Math Destruction.” *Discover*, vol. 37, no. 8, Oct. 2016, p. 50-55, *EBSCOhost*, search.ebscohost.com/login.aspx?direct=true&db=f5h&AN=117469935.

Roose, Kevin. “How Facebook's Data Sharing Went from Feature to Bug.” *The New York Times*, 19 Mar. 2018, www.nytimes.com/2018/03/19/technology/facebook-data-sharing.html.

Singer, Natasha. “What You Don't Know About How Facebook Uses Your Data.” *The New York Times*, 11 Apr. 2018, www.nytimes.com/2018/04/11/technology/facebook-privacy-hearings.html.

University of Copenhagen. “Data Power? What the Foucault!” *ETHOS*, 17 Nov. 2016, ethos.itu.dk/2016/10/17/data-power-what-the-foucault/.