Uncertainty and Desire: Big Data Surveillance and Digital Citizenship

Jamie Duncan

Abstract

This synthesis presents the works of several key scholars in the field of critical data studies to explore how the practice of Big Data surveillance is redrafting possibilities for digital citizenship and contemporary democracy. Through a critical analysis of five works, I establish that the distinction between public and private surveillance is becoming ever-more unclear. In this light, I explore how the technology interfaces with inequality, transparency, and online participation, calling into question the validity, ethics and consequences of Big Data surveillance. I discuss how these phenomena construct ambivalent digital citizens who are both willing participants and fearful subjects in their online activities. I conclude that transparency, democratic accountability and novel forms of activism and policymaking are needed to counter the use of Big Data as a tool of political control and manipulation.

Keywords: big data; politics; surveillance; digital citizenship; inequality

Uncertainty and Desire: Big Data Surveillance and Digital Citizenship

As Big Data analysis continues to be celebrated as an emancipatory use of technology, academic efforts to describe the socio-political implications and to mitigate risks are well underway. This review will explore how social practices inform Big Data analysis by consulting works from the following researchers: a team of academics who offer a policy framework designed for governments to improve Big Data policy (Broeders et al., 2017); Lyon (2017), who parses language used in surveillance discourse, offering a political-sociological description of participatory digital citizenship and contemporary surveillance; boyd and Crawford (2012), who call for critical evaluations of Big Data using a socio-technical lens; Brayne (2017), who provides one of the first practical studies of the technology applied to police work, focusing on inequality; and Ruppert, Isin and Bigo (2017) who discuss the socio-political implications of contemporary data practices for engaged citizenship. While the term Big Data itself is avoided by Lyon, and used sparingly by Ruppert et al, data analysis characterized by large, varied
data sets at high speeds is becoming ever-more commonplace. I argue that Big Data surveillance is changing possibilities for citizens' interactions with states and markets, and that these developments require creative policy interventions and new understandings of citizenship. I will describe these shifts and present emergent policy challenges for the three key themes of social inequality, governmental and corporate transparency and online participation. Social inequality is analyzed in the context of Big Data surveillance. Next, transparency is discussed in the context of black-boxed and proprietary algorithms and datasets as they relate to state and market cooperation. Finally, the paradoxical role of the digital citizen as a participant in mass-surveillance will be discussed.

**On Inequalities and Errors**

The potential to increase inequality and hypostatize falsehoods using advanced data analysis are examples of how Big Data interacts directly with existing social policy challenges. Some tout the technology as liberating, believing algorithmic methods produce consistent and objective results, rectifying human fallibility (Brayne, 2017; boyd and Crawford, 2012). This logic fails to account for inextricable socio-political influences on data collection, calculation and analysis. The technology presents opportunities for businesses, governments and everyday people in its potential to limit corporate and financial fraud, make mundane tasks like finding goods and services easier and empowering everyday people as researchers through access to open data sources; however, it also has potential to entrench existing inequalities and create new ones (Broeders et al., 2017; boyd & Crawford, 2012). Brayne’s (2017) study of Big Data in the everyday operations of the Los Angeles Police Department (LAPD) uncovers how trends like the criminalization of low-income, minority neighbourhoods are amplified through Big Data analysis. Among the first “on-the-ground accounts” (p. 978) of Big Data in public institutions, Brayne’s study provides examples to back up many of boyd and Crawford’s (2012) speculations using interviews and observations collected over the course of more than two years of fieldwork (Brayne, 2017). Faith in Big Data is high and outputs are often accepted without much critical thinking about how they were obtained (Brayne, 2017; Broeders et al., 2017). Moreover, social inequalities reflecting past and present social relations are represented in existing statistics and classifications (Brayne, 2017). Numbers that are flawed, incomplete, and informed by historical bias are being used to train algorithms, consequently promoting behaviours that reinforce the social status quo. In Los Angeles, actuarial algorithms assign higher risk scores to individuals and neighbourhoods that are associated with lower incomes or racialized populations, resulting in more interactions with police, which further inflates risk scores (Brayne, 2017). The amplification of inequality has real implications, resulting in disproportionate arrests and detentions (Brayne, 2017). This type of self-authentication discredits the idea that recursive algorithms trained with “just the numbers” somehow offer objective results. The potential for predictive policing to perpetuate inequality, promoting racist and classist norms supports Broeders et al.’s (2017) identification of the need to regulate not just the origins of data, but also calculation and analysis. Each of these stages present opportunities to infuse human or organizational
bias into the process and skew outcomes (Broeders et al, 2017).

In addition to exacerbating inequality, boyd and Crawford (2012) predicted that new digital divides will emerge between data owners and producers, preventing the public from accessing the benefits of technology and empowering private companies and security agencies to predict and influence the actions of citizens. This has become a reality; Big Data offers a new method for the wealthy to manipulate political outcomes. Ruppert et al (2017) cite Cambridge Analytica, a company that leveraged Big Data to target ads and influence results in the 2016 American election as well as the UK’s Brexit referendum. Furthermore, Brayne’s (2017) study demonstrates how this new divide affects power dynamics between police officers and citizens. Both cases demonstrate the emergent relationship between data owners and producers. To defensively regulate against potential abuses of this power, boyd and Crawford (2012) argue that policy makers must rely on social science skills—in addition to technical ones—to reflect on and account for biases and critically engage with positivist and determinist discourses. For citizens, actions taken in the digital realm are affecting democratic processes in the physical world more directly than ever before. Claiming digital rights in such transversal contexts is not only justified but necessary (Ruppert et al, 2017). This emergent data politics presents opportunities for novel acts of digital citizenship and resistance to state and corporate control through claims to individual agency and data ownership (Ruppert et al, 2017; Lyon, 2017). Nonetheless, a significant challenge to empowering digital citizenship is the high level of secrecy maintained by governments and private actors.

**Black Boxes and Big Profits**

It is well established that states and corporations collaboratively engage in Big Data surveillance practices like predictive policing and social media intelligence gathering (Brayne 2017; Lyon 2017; Broeders, 2017). However, not enough is known about how they engage in these practices. Ultimately, this lack of clarity reifies inequalities between data owners and producers. Transparency during the collection of data, application of algorithmic methods, and analysis of outcomes is fundamental to fostering trust in Big Data technology (Broeders et al. 2017). States and markets have distinct reasons for black-boxing surveillance methods and algorithms but they also cooperate because their interests often intersect. Ruppert et al. (2017) believe that the enumeration of populations and control of knowledge are core aspects of state control. Conversely, Broeders et al. (2017) explain that state actors charged with upholding the security of their populations need secrecy to ensure operational integrity and public safety. Lyon (2017) complicates this by mentioning some ugly truths like those unveiled by Edward Snowden that were hidden by states in the name of national security. Together, these points uncover a conflictual ethics of state power characterized by tensions between the maintenance of control and the defence of democratic values. In the exceptional context of national security, the need for control allows states to obfuscate their actions in ways not typically tolerated by the public.

Much more simply, corporations hide their algorithms to protect profits, seeking a different brand of power. There is a lot of money to be made in the industry of surveillance (Broeders et al. 2017;
Lyon 2017; Ruppert et al 2017). Private actors are heavily involved in data collection and brokering information to those who are willing to pay, including government security agencies (Lyon, 2017). Brayne (2017) specifies that security agencies often bolster public data by purchasing private data, which is subsequently run through private algorithms sold as a service. Lyon (2017) and Broeders et al. (2017) also describe how states often just take the information they want, using social media intelligence and backdoors into corporate databases. This private data allows security agencies to make intimate personal connections that once required a warrant (Brayne, 2017). Big Data is blurring the difference between public and private surveillance in this way, posing big questions about individual rights, accountability and data ownership. Broeders et al. (2017) describe how social media intelligence and the availability of private data have changed the relationship between states and markets. With respect to this relationship, Lyon (2017) suggests that both benefit from the proliferation of data consumption and production as it contributes to their distinct organizational missions. boyd and Crawford (2012), working at Microsoft Research, also explain that industry threatens to eclipse traditional research institutions through proprietary access to valuable datasets, changing how research will be understood and practiced. Finally, Ruppert et al (2017) frame these trends as a reconfiguration of knowledge structures constituting an emergence of new politics and peoples, thereby changing the rights and responsibilities of citizenship. Creative policy and activist interventions are needed to protect public interests and ensure the validity and reliability of processes and outcomes related to emerging applications of Big Data analysis.

**Uncertainty, Desire and the Ambivalent Digital Citizen**

Bits and bytes dredged up from the internet are clearly not as innocuous as they seem. Through database centralization, supposedly anonymous information can be assembled to become not just personally identifiable, but threatening to legally enshrined civil liberties. Lyon (2017) posits that surveillance has become so normalized that significant portions of the population actively and knowingly participate to the benefit of both states and corporations. This can be explained by constructing people both as objects of surveillance and subjects active in shaping its terms and outcomes (Ruppert et al., 2017; Lyon, 2017). Despite the seeming banality of data collection, states and corporations are collaborating to monitor and influence digital citizens in ways that threaten individual privacy and freedom by predicting criminality, monitoring online interactions and manipulating democratic outcomes (Broeders et al., 2017; Brayne, 2017; Lyon, 2017; Ruppert et al., 2017). As a socially embedded and politically informed practice, Big Data analysis requires significant oversight to ensure benefits to the public are maximized and risks are mitigated through policy and advocacy. Similar to Ruppert et al (2017), boyd and Crawford (2012) signal that Big Data presents a dangerous shift in the overall meaning of knowledge and that social identities and norms will be altered by these new ways of thinking and doing. The ontological and epistemological threat they identify is characterized by the belief, in some, that numbers speak to truth and offer all the context needed to understand reality (boyd & Crawford,
This logic is imbricated with an almost religious faith in technological emancipation and fails to account for the socially mediated nature of data collection, calculation and analysis. As a way of knowing, this form of positivism threatens to minimize the importance of human discretion, intuition, and common sense, which are necessary for addressing the socially mediated nature of quantitative methods. For this reason, boyd and Crawford (2012) argue that social scientists be consulted in the creation and implementation of algorithmic technologies to ensure ethical and accurate engagement with Big Data as a quantitative tool and not an erroneous answer to uncertainty.

Considering the potential for abuse, as well as the political-economic implications of these surveillance methods, advocating for transparency is key to ensuring that advanced data analysis is used ethically and effectively (Broeders et al., 2017). Ruppert et al (2017) argue that better understanding how data is politically situated in the everyday lives of people on- and offline is a precondition to effectively exercising digital rights. Lyon’s (2017) view that internet users are subject-participants in surveillance means that despite the seemingly inescapable quality of state and corporate efforts to track and enumerate users, people negotiate with power and affect the processes of collection, calculation, and analysis of their information as digital subjects, and identify sites for empowerment and activism. He also describes how digital citizens alter their online activities in response to new information on how they are being watched (Lyon, 2017). Being educated on how information is collected, stored, analyzed and acted upon is therefore fundamental to meaningful online participation and effective dialogue between citizens, states and markets (Ruppert et al., 2017). Despite this potentiality, Broeders et al. (2017) indicate that advocating for absolute transparency is not necessarily the most effective way of negotiating with power. They propose a system of layered accountability, whereby a watchdog agency verifies the origins of data, analyzes methodologies and establishes the validity of outcomes towards the end of safeguarding democratic rights and freedoms (Broeders et al, 2017). This would hypothetically occur without compromising national security or the integrity of corporate intellectual property. The public would be informed on surveillance practices in a redacted report offering limited accountability (Broeders et al, 2017). While this does offer opportunities to reign in and regulate corporate algorithms, it does not fully address the possibilities for digital citizenship described by Lyon (2017) and Ruppert et al (2017). Governments would still be effectively accountable to themselves and the public would remain relatively ignorant. A high level of transparency is required for educated digital decision-making and for state and corporate accountability.

Despite being conscious of their ignorance, people still actively participate in their own surveillance. Uncertainties surrounding Big Data surveillance are more complex than just the increasingly hazy difference between state and market surveillance. The lines between awareness and ignorance, desire and fear, as well as complicity and resistance are unclear. One cannot be sure of the consequences of one’s own actions. Ambivalent identities and enactments of digital citizenship emerge from this uncertainty. Lyon (2017) and Ruppert et al (2017) explain practices of participatory surveillance as manifestations of desire and agency. The desire to be seen and heard by others is a primary driver of
contemporary surveillance. This is particularly true among younger people, some of whom perceive a social reality in which existence itself hinges upon online presence (boyd, 2014 in Lyon, 2017). In a society where physical modes of being are increasingly mediated by digital ones, understanding ourselves as empowered digital citizens becomes necessary to enacting agency and addressing the ambiguities in modern data politics and Big Data surveillance. In these ways, we participate in shaping data as it shapes us (boyd & Crawford, 2012). Lyon (2017) explains this tension between desire and uncertainty through the concept of “surveillance imaginaries”—individual behaviours and understandings of surveillance that vary greatly between individuals (p. 834). Lyon (2017) identifies a tension between faith in corporations as benevolent monitors and uncertainty surrounding the undesirable implications of the mass-surveillance publics know is occurring. These seemingly opposite perspectives co-exist, a person can be both afraid and complicit. In their attempt to explain the paradoxical nature of digital citizenship, Ruppert et al (2017) suggest that despite the various ways people demonstrate agency, data “captures [and] colonizes minds, souls, bodies, and spaces” (p. 5). Through informational affect, the norms of “data’s empire” promote a desire for the paradoxical agency of being both the watched and the watcher in the contexts of hyper-individualism and performed authenticity (Ruppert et al, 2017, p. 5; Lyon, 2017). The desires of digital citizens are thus constructed and socialized in ways that form subjective understandings of surveillance and online participation, which often result in the generation of information that can be bought, sold, and leveraged for control. The meaning of citizenship is changing as interactions—and some say existence—are increasingly dependent on online presence and the generation of online content. Inherent in these evolving relations between governments, corporations and everyday people are social and political risks that pose unique policy challenges related to transparency, inequality, and civil liberties.

**Conclusion**

The seemingly inevitable ubiquity of technologies like Big Data in commerce, politics, and social interaction will continue to create new opportunities for oppression and activism in our lives as digital citizens. We can expect policy interventions like those advocated for by Broeders et al. (2017) to partially mediate this process, however, significant tensions remain between and within the interests and identities of citizens, corporations and the state. In this synthesis of works by several leading scholars I have argued that the advent of Big Data surveillance is changing possibilities for interactions between the private sector, governments and citizens, requiring new understandings of citizenship and novel policy approaches to address the implications for issues like social and economic inequality, corporate and governmental transparency and online participation. I expect that data rights and digital citizenship will be topics of heated policy debate and political activism in coming years as relations between states, markets and citizens continue to be renegotiated.
References


