Designing for democracy
Bulk data and authoritarianism
Robbins, Scott; Henschke, Adam

Publication date
2017

Document Version
Final published version

Published in
Surveillance and Society

Citation (APA)

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.
Abstract

Transparency is important for liberal democracies; however, the value of transparency is difficult to articulate. In this article we articulate transparency as an instrumental value for providing what we call ensurance and assurance to liberal democratic citizens. Ensurance refers to the property of liberal democracies which prevents it from sliding into authoritarianism and assurance is the property whereby citizens are assured that ensurance exists. Looking at the rise of bulk data collection and use afforded by information communication technologies, this paper focuses on the way that technologies disrupt relations between the state and its citizens, and suggests Value Sensitive Design as a methodology to protect key aspects of liberal democracies.

Bulk data collection makes the achieving of ensurance and assurance more difficult due to two types of opacity which arise as a result of the practice: technical opacity—the difficulty for citizens to understand the technology behind bulk data collection; and, algorithmic opacity—opacity which results from properties inherent to algorithms which guide the collection and processing of bulk data. Design requirements will be suggested to respond to the disruptions caused by ICTs between liberal democracies and their citizens which threaten the necessary value for liberal democracies of representativeness.

1. Introduction: Disrupting Relations between Citizens and the State

Following the revelations by Edward Snowden about widespread state sponsored surveillance programs (Greenwald 2014; Harding 2014) some fear that liberal democracies are at risk of descending into an authoritarianism. “Obviously, the United States is not now a police state. But given the extent of this invasion of people’s privacy, we do have the full electronic and legislative infrastructure of such a state…These powers are extremely dangerous” (Ellsberg 2013). The revelation of comprehensive government surveillance programs implies that liberal democracies are about to become authoritarian police states.

Despite many legitimate concerns about state overreach and worries about the national security organs in liberal democratic states, the US, UK, and other similar countries remain worlds apart from somewhere like North Korea: perhaps people like Daniel Ellsberg are worried over nothing? However, looking at modern surveillance technologies suggests that there is a disruption of relations between citizen and the state. Snowden’s revelations shed light on the notion that a fundamental shift had occurred between liberal democratic states and their citizens—in the US, for example, the National Security Agency (NSA) had
access to vast amounts of information about its citizens, while the citizens did not know the extent of this. “These bureaucratic ways of using our information have palpable effects on our lives because people use our dossiers to make important decisions about us to which we are not always privy” (Solove 2004: 9). This is highly important as the relation between liberal democratic states and their citizens is absolutely central to them being liberal democracies—should such a state cease representing those citizens, then it can no longer properly be called a liberal democracy.

The information communication technologies (ICTs) that enable comprehensive surveillance disrupt the state/citizen relation due to their ‘opacity’; what Snowden showed was a ‘revelation’ not so much because it publicized specifics about state surveillance programs, but because these programs were being done in the name of these state’s citizens without their knowledge. “While the government, via surveillance, knows more and more about what its citizens are doing, its citizens know less and less about what their government is doing... Democracy requires accountability and consent of the governed, which is only possible if citizens know what is done in their name” (Greenwald 2014: 208–9). We ask if a state can be representative of the will of its citizens if those citizens do not know what the state is doing.

‘Representativeness’ is core to liberal democracies. Here, “public authorities are bound by their own rules and can only exercise their powers in a lawful way. All powers must derive from the constitution… [implying] the important fact that the government is accountable and that its actions must be controllable, and thus transparent” (Gutwirth and de Hert 2006: 64). This representativeness has elements of legitimacy and control in it. First, on the social contract view, the legitimacy of the government is dependent on the state actually representing the view of those it governs. Second, for our purposes here, the governed need to know what is being done in their name, such that they can remove legitimacy should it become apparent that the state is not actually representing the will of the governed (Altman and Wellman 2009: 3–6). The strong public condemnation of the US surveillance programs following Snowden is evidence that many US citizens did not know what was being done in their name. Bulk, open ended surveillance programs are problematic specifically for this—the information gathered is potentially limitless in its use and who can use it.

This points to the deeper disruption being caused by ICTs. A recent article in the Scientific American asked whether democracy will survive big data and artificial intelligence (Helbing et al. 2017). The authors’ concerns stem from the ways that increasingly sophisticated ICTs can be used by states and private companies in ways that citizens and consumers are largely ignorant of. What we are suggesting here is that ICTs are disrupting the relations between the citizen and the state by giving the state unparalleled access to information about its citizens, while the citizens are not comparably informed about what the state is doing or what it knows. This is what we can call an ‘informational deficit’, where the state’s knowledge about its citizens substantially surpasses what the citizens know about the state. While there has always been some informational deficit between what a state does and what its citizens know, the worry here is that the new technologies provide so much more information about its people, without a corresponding increase in the citizens knowing about the state. Furthermore, these technologies are increasingly complex to the point of being opaque to citizens’ understanding. Not only are the ICTs difficult to understand due to their complexity, but artificial intelligence (AI) used to process bulk data collected by the state can be inherently opaque—even to operators of the technology. The value of representation must be realized in the design of these ICT systems.

### 2. Designing For Representation: Ensuring and Assuring the Will of the Citizens

Liberal democracies are by definition not authoritarian. “The original impulse of the liberal tradition, found in Locke and Kant, is the idea of the moral sovereignty of each individual. It implies limitations on the ways in which the state can legitimately restrict the liberty of individuals even though it must be granted a monopoly of force in order to serve their collective interests and preserve the peace among them” (Nagel
In the liberal tradition, we forgo certain rights in order to have collective goods such as security. Core aspects of liberal democracies protect against authoritarianism, they have processes to ensure and assure us that they are not authoritarian. On ensurance, a state cannot become authoritarian. And on assurance, the citizens know the state is not authoritarian. We suggest that the ICT systems that support bulk-data collection can be designed with processes that ensure and assure a state’s citizens for the sake of representation.

The methodology we use is a specific articulation of Value Sensitive Design (VSD). VSD starts with the idea “that aims at making moral values part of technological design, research and development. It assumes that human values, norms, moral considerations can be imparted to the things we make and use” (van den Hoven 2007: 67). This is particularly relevant when looking at the relation between citizens and the state, and how those relations can be disrupted by technologies. “If we want our information technology - and the use that is made of it—to be just, fair and safe, we must see to it that it inherits our good intentions. Moreover it must be seen to have those properties, we must be able to demonstrate that they possess these morally desirable features” (van den Hoven 2007). What follows is a brief conceptual investigation (as per VSD) of the value of representation and end-norms to ensure and assure that liberal democracies to not become authoritarian.

One step in actively designing for a particular value is to specify that value—here we are particularly concerned with the value of representation in liberal democracies. We suggest that a focus on two design elements can go some way to translating representation into the design requirements of state surveillance technologies. Here, we consider that surveillance (and other state based bulk data collection programs) should be designed such that they ensure the will of the citizens is represented, and that the citizens are assured that the state’s actions are representative of their will. This draws from the notion of a values hierarchy in VSD, in which the design of a ICT system brings in ‘end-norms’ for the sake of the ultimate value being designed for. “End-norms in design then may refer to properties, attributes or capabilities that the designed artefact should possess” (van de Poel 2013: 258).

The two end-norms that we consider of prime importance to bulk data and representation are to ensure and assure that the state represents the will of its citizens. ‘Ensurance’ is the attribute of a system in which the design features ensure that an end is either being achieved or not frustrated by technologies. To ensure that a surveillance program is neither being used in the citizen’s name without their knowledge or is being used against the state’s own citizens, the citizens or their representatives must be informed about how bulk data is collected, how it is being used and who has access to it and so on. Here, insofar as bulk data about citizens is collected and potentially used against a state’s own citizens in secret, then ensurance has not been met, and representation has been undermined. Reducing informational deficits can ensure that there is no slide into authoritarianism. That is, if the citizens don’t agree with what’s being done in their name, then they can withdraw support for the state. This capacity to withdraw support is core to representative democracies. However, in cases of informational deficits the citizens don’t know what’s being done, so can’t know when to withdraw support.

The second end-norm is concerned with assuring a state’s citizens that the state is representing their will. Here, the process is about the confidence that citizens have that their will is represented. A “[g]overnment therefore has to explain through the media the rationale for the strategy it is following and convey a sense of where and why it is balancing the benefits from additional security with all the costs of providing it” (Omand 2010: 18). An informational deficit between the citizen and the state is concerning because such

---

1 Here, we are agnostic about security as an intrinsic good, or an instrumental good that protects other collective goods.
2 Here, ‘representative’ refers to something like the US Foreign Intelligence Surveillance Act (FISA) courts in which oversight is achieved through those acting on behalf of the citizens.
deficits can cause changes in citizen’s behavior. A Pew Research Center poll showed that a quarter of Americans have changed their behavior with regard to technology due to US government surveillance (Gao 2015). This is the so-called “chilling effect”—when governmental regulation and policy not directed at certain activities deters individuals from carrying out protected activities. Currently, bulk data collection in the name of national security is directed at terrorists and particularly bad criminals. However, if being concerned about government surveillance, a citizen is deterred from participating in legitimate political organization and activism, then the chilling effect has taken place. The Pew research Centre report shows, this “chilling” has in fact effected the behavior of a quarter of Americans. What is interesting is that this effect takes place even if policies to protect citizen’s privacy exist. This is because citizens don’t have the knowledge needed to feel assured: on this, good policy must not only exist, but be known by the citizen to exist. informational deficits can keep the chilling effect in place.

3. The Instrumental Value of Transparency

Transparency is the solution to informational deficits. However, transparency is not a value in itself; rather, it is instrumental for realizing other values such as representation that are important. Transparency can be “an ethically “enabling” or “impairing” factor (Turilli and Floridi 2009). Thinking about transparency in this way prevents radical approaches in which the only option is full disclosure—which would undermine the values (e.g. safety and security) which the government is setup to realize. The goal for governments is to use transparency to enable ethical values which are lacking due to the informational deficits outlined above.

Transparency can be used to ensure that privacy is not being overridden without justification and authorization and to assure citizens that there are appropriate policies in place. However, the simple fact that this information is available does not mean that assurance has been realized. The process of effectively disclosing this process is important:

Information transparency should disclose not only information but also details about how such information has been produced. Such details are a necessary condition for verifying the consistency between the ethical principles endorsed at the time of producing information and the ethical principles that information transparency should enable. (Turilli and Floridi 2009: 109)

In order for transparency to fulfil its instrumental value, information needs to be disclosed to the public in a way that can realize the value of an assured citizenry. For instance, the processes by which this information is chosen, prepared and redacted may also need to be disclosed. Before we can articulate how to realize representation through transparency in the design of ICTs which enable bulk data collection we must understand how these ICTs reduce transparency.

4. Bulk Data Collection and Opacity

Bulk data collection has made the quest for transparency a moving target. This is so for two reasons: first, the technology (both infrastructure and code) behind bulk data collection is difficult for the public to understand; and second, some algorithms used to process bulk data are intrinsically opaque due to properties of contemporary approaches to AI and machine learning. Both of these reasons make the regulation as well as the assurance and insurance discussed earlier difficult to realize.

---

3 For more discussion on opacity and machines see Burrell (2016).
4.1 Technical Opacity

While debating net neutrality—the idea that internet service providers should treat all content equally—Senator Ted Stevens famously said that the internet is “a series of tubes” (Belson 2006). While there is some debate about how accurate that metaphor is, it represents an opacity which has serious consequences for the regulation of the internet. Senator Ted Stevens, despite his technical illiteracy, gets a vote on important legislation—and people who are equally technically illiterate vote him into office.

With respect to bulk data collection, for example, much has been discussed about the NSA tapping the backbone of the internet (Greenwald 2014; Kravets 2013). The NSA has partnered up with telecommunications companies, and other liberal democracies (the Five Eyes4), to place fiber optic cable splitters on the cables that serve as the backbone of the internet. This, effectively, puts a “tap” on the internet. The information which flows through this tap must be filtered5 and stored on government servers. To properly understand what this means to citizens, one would have to understand how much of the internet’s traffic flows through these taps, what filters are in place to ensure data is not collected from citizens, and how much ‘incidental’ citizen data is collected. A further point which is important but will not be explored here is what institutional and legal arrangements are made to prevent governments from bypassing restrictions on collecting citizen data by simply obtaining the data from other countries.6

While there have been speculation and educated guesses about how much data flows through these taps7, the filters used and how successful they are at both preventing citizen data collection and preventing terrorist attacks is unknown to the public. Without an understanding of how much of their privacy is being “incidentally” invaded, and how effective this technological solution is with regard to terrorism there is no way that the public can be assured that the state is not overreaching. Furthermore, citizens cannot consent to such a program without having some degree of knowledge about how effective it is. For example, if it were true that these technologies would have prevented 9/11 (a hypothetical that is widely disputed)8, then citizens may conclude that it is worth it to have this program even if it has the potential to invade their privacy.

4.2 Algorithmic Opacity

Artificial Intelligence (AI) uses machine learning algorithms which are amazing at categorizing things. For example, Google image search is able to quite accurately categorize images.9 And now there are proposals coming both form academia and private companies for how to use AI to combat terrorism (Aviv 2009; Frenkel 2017), discover illegal immigrants, catch tax evaders (Hemberg et al. 2016), etc. Governments are feeding algorithms bulk collected data so that algorithms can make decisions about us.

AI methods are being increasingly employed to enhance a system’s ability to reach decisions about large data sets. AlphaGo, developed by Google uses a method called deep learning to “learn” how to play the game Go. When AlphaGo makes a move, not even the programmers understand why it made the move that

---

4 ‘Five Eyes’ is an intelligence sharing agreement between the US, UK, Canada, New Zealand, and Australia.
5 By US law there must be a procedure in place for minimizing US personal data. The FISA court approves these procedures once a year.
6 The FISA Courts in the US, for example, are particularly attentive to this citizen/non-citizen (or, more precisely a US person/non-US person) distinction, which points to a third form of opacity, legal opacity. However, we do not have space to cover legal opacity in this paper.
7 https://sniffmap.telcomap.org/ tries to show how much data the NSA and its partners intercept.
8 Former FBI chief Robert Mueller claims that bulk data collection would have prevented 9/11 (Roberts 2013) while CNN national security analyst and journalist Peter Bergen forcefully argues against this idea (Bergen 2013). An in depth read about this can be found in a 2015 New Yorker article (Schwartz 2015).
9 There have been some embarrassing mistakes however. See BBC News (2015).
it did. This is algorithmic opacity—opacity which is a result of the properties of an algorithm itself. This makes the decisions made by AI algorithms opaque to even those who are technically literate.

5. Restoring Representation: Transparency by Design

Having established that transparency plays an important role in preventing a government’s slide into authoritarianism by instrumentally supporting assurance and assurance, it is necessary that governments do what they can to limit technical and algorithmic opacity. “For this, the state would have to provide an appropriate regulatory framework, which ensures that technologies are designed and used in ways that are compatible with democracy… Individuals would then be able to decide who can use their information, for what purpose and for how long” (Helbing et al. 2017).

What follows is a brief technical investigation (as per VSD) of the challenges and corresponding design requirements associated with the systems involved in government bulk collection of meta-data. This is not a detailed, exhaustive solution to the problem of opacity. It is an important step towards the ideal of a representative, transparent, and legitimate liberal democracy.

5.1 Technical Transparency

Making the technical aspects of bulk data collection more transparent to technically illiterate public is important for assuring that governments are not overreaching. Non-Governmental organizations in the US like the Electronic Frontier Foundation (EFF)\footnote{See https://www.eff.org/} have gone some way to doing that by providing info graphics and easy to read descriptions of how governments bulk collect data. However, the stance of the EFF is decidedly directed at problems with the government’s bulk data collecting programs. The government should make a concerted effort to take the lead in explaining the what, when, how, and why to keep the public assured that there is no government overreach.

A design requirement which would go some way in realizing the value of transparency would be to audit bulk collected data and remove citizen’s data. This solution could involve human auditors, or, because there is so much data, algorithms. This would enable the ability to report on how much incidental data is collected and the processes in place to remove that data. While citizens would not necessarily understand the technical processes behind bulk collection, knowledge of the results of these processes would help tremendously. This should include the efficacy of these processes. How good are these processes at preventing terrorism? The US government at least has been silent about this.\footnote{In a very recent congressional hearing, the NSA attempted to give examples of the success of bulk data collection—none of which clearly showed that this data helped to prevent attacks in liberal democratic countries (Savage 2017).} Without this knowledge, citizens cannot begin to balance the values of privacy and security. While making transparent the details of how exactly a specific terrorist plot was prevented may compromise security, general reporting on the success of bulk collection programs is necessary to realize transparency.

To sum up, two specific requirements result if the government were to design these technologies which realize the value of transparency—which is instrumental to both assurance and assurance: first, the ICT must include the capability of auditing the system for incidentally collected data associated with citizens; and second, there must be the ability to report on its success in preventing terrorism. Institutional and legal arrangements should be made to distribute this information to the public (in a way that does not compromise the success of the ICT). Without these processes the will of the public is not effectively represented.
5.2 Algorithmic Transparency
Making the decisions of AI algorithms transparent is a hot computer science topic. Researchers and companies seem to understand that decisions made by algorithms will not be tolerated if we cannot understand them. No one wants to be prevented from getting on a plane because an algorithm put them on the No-Fly list without an explanation. The nature of some of these algorithms (e.g. deep learning) make a solution to opacity extremely difficult and we should not expect that this will be accomplished anytime soon.

The solution, therefore, is to use such algorithms for specific situations in which it is acceptable to not have an explanation or to supplement the decision of the algorithm with human oversight. Placing someone on the No-Fly list, for example should not be solely decided on the basis of an algorithm which can offer no explanation. A restriction of one’s rights is a moral decision and only a human being can accept the moral responsibility which comes along with such a decision.

The design requirement which comes out of this will help realize the value of ensurance. Ensurance is realized in this situation if the government is prevented from using these kinds of algorithms for moral decision making. Policy should be written to show that this is the case and this policy should be made public so that citizens are assured. Only with such assurance will representation be realized.

***

This brief technical investigation helps to move liberal democracies closer to realizing the instrumental value of transparency. Transparency is important for ensurance that liberal democracies cannot become authoritarian. Transparency is also important for an assured public—a public confident that the property of ensurance has been met. VSD is key to responding to the disruptions caused by ICTs between liberal democratic states and their citizens. By specifying representation as a value and highlighting connections between transparency and technology we can design ICTs for democracy.

Acknowledgement
Funded by the European Research Council grant M44 A33 “Global Terrorism and Collective Moral Responsibility: Redesigning Military, Police and Intelligence Institutions in Liberal Democracies”.

References

12 See for example a recent PEW Research Center report (Rainie and Anderson 2017).
13 For more on this discussion see Bryson (2010) and Johnson (2006).


