

## [A Neurophilosophy of Big Data & Civil Liberties, and the Need for a New Social Contract](#)

By Nayef Al-Rodhan

May 21, 2020



*This is post three in a short-term series by Prof Nayef Al-Rodhan titled “Neurophilosophy of Governance, Power and Transformative Innovations.” This series provides neurophilosophical perspectives and multi-disciplinary analyses on topics related to power and political institutions, as well as on a series of contemporary transformative technologies and their disruptive nature. The goal is to inspire innovative intellectual reflections and to advance novel policy considerations.*

Greek mathematician and philosopher Pythagoras (6<sup>th</sup> century BC) is believed to have likened everything to numbers. In the writings of Aristoxenus (4<sup>th</sup> century BC), Pythagoras is said to have argued that [numbers explain and control things](#). Over twenty-six centuries later, in the age of Big Data, this appears as a particularly prescient claim.

The large amounts of data sets generated and multiplied today on an hourly basis raise important philosophical questions for the future of governance and the foundation of the social contract as we know it. These questions are new because they create entirely new forms of power imbalances in society.

There are numerous aspects of Big Data already being explored from ethical and philosophical perspectives, including regulatory challenges and [epistemological implications](#) of the ‘data revolution’. Big Data’s implications are poised to reach most aspects of our lives, and this includes our rights, liberties and ultimately, the social contract. This article provides a neurophilosophical perspective on Big Data, its relations to human nature and the implications for governance.

### **Big Data – how much data?**

While it is difficult to pinpoint the total amount of data ever produced (taking into account that it grows each second), even rough approximates are enough to paint an astonishing picture. It is estimated that from [the beginning of recorded data to 2003](#), about 5 billion gigabytes were generated. By 2015, 5 billion of gigabytes were being generated every 10 seconds. [Between 2016-2018](#), about 90% of the data in the world was generated. In 2018, 2.5 quintillion bytes of data were created each day, a pace that is only increasing with the Internet of Things, meaning with smart devices and machines connected to the internet. For example, [each connected car](#) generates about 4 terabytes

every day. By 2025, estimates place the generation of data at 463 exabytes per day globally. The increase of data means that we also need to operate with [data scales](#) that require new data literacy.

[Data is measured](#) in bits and bytes, with 1 bit containing a value of 0 or 1, and 8 bits make a byte. Kilobytes have a value of 1000 bytes, megabytes of  $1000^2$  bytes, gigabytes  $1000^3$  bytes, terabytes  $1000^4$  bytes, petabytes  $1000^5$  bytes, exabytes  $1000^6$  bytes and zettabytes  $1000^7$  bytes.

The yottabyte, with a capacity of  $1000^8$  bytes is currently the largest storage term in the International System of Units. While the existence of this data seems somewhat of an 'ethereal' reality, it is of course reliant on physical infrastructure including servers and data centers, such as the one built by the US National Security Agency in Utah (the Utah Data Centre). Furthermore, not only does data rely on physical infrastructure, the space needed for digital files is anything but small. For example, a [zettabyte](#) would need to fill approximately 1000 datacenters, or 1/5 of Manhattan.

However, the unique qualities of Big Data do not reflect in volume alone. Governments, industries and academics have experience producing and working with large data sets, for instance in the context of national censuses. The data compiled in the context of such demographic studies is usually generated through inflexible methods (for example, standard questions that cannot be tweaked later), and, given the significant effort, this exercise is usually conducted every five or ten years.

In marked contrast, [Big Data](#) is generated *continuously*, is flexible and fine-grained in scope. 'Big Data' therefore is not strictly characterized by size, but also by the range of [computational methods](#) that are used 'to group and analyze data sets'. [Rob Kitchin](#) summarizes the characteristics of Big Data as data sets that are high in *volume*, high in *velocity* – being created in real time, diverse in *variety* – being structured and unstructured, *exhaustive* in scope – meaning striving to capture entire population groups or systems, fine-grained in *resolution*, *relational* in nature – meaning that different data sets can be conjoined, and *flexible* – which implies extensionality (can easily add new fields) and scalability (can expand in size).

These fundamental features of Big Data (high volume, velocity, variety, exhaustivity, resolution, relationality and flexibility) have implications that go far beyond the commercial sector. Big Data contributes to new methods of knowledge production and legitimizes new methods of engaging with – and trusting – mathematical formulae.

In a 2008 [Wired piece](#), Chris Anderson synthesized the characteristics of this new *empiricist mode* of knowledge production. Across different disciplines, from physics to biology and genetics, the established scientific script of 'hypothesize – model – test' seems increasingly obsolete, as data is entrusted to explain everything. And that means that what we are starting to accept as valid knowledge need not be derived from *theories* of human behavior, or scientific models (e.g. the Newtonian model), but from other methods.

In the petabytes era, "[correlation is enough](#)", Anderson wrote: the patterns and relationships contained within Big Data produce meaningful knowledge about highly complex phenomena without a need for hypotheses or even coherent models or unified theories. This is exactly how a lot of algorithmic data is processed at Google, for instance, without inquiring about underlying reasons or without hypotheses about what the results may show:

"Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves."

This approach is also increasingly adopted in scientific fields, including in quantum physics, biology, in the field of neuroscience or other scientific domains such as epigenetics. The sequencing of the human DNA (with the first results published in *Nature* and in *Science* in 2001) led later on to other discoveries, including of new species of bacteria – all thanks to supercomputers and high-speed sequencers. These findings can then be correlated or analyzed together with other findings and advance new knowledge. In short, [Big Data](#) can look at vast amounts of information, and provide full resolution, without the need for a priory theory.

The hype created around Big Data has also created many – at times, unrealistic – expectations that by crunching huge sets of data, we will be able to tackle virtually all international crises, from immediate and recurrent ones, such as [pandemics](#), to existential and long-term challenges such as [climate change](#). However, as demonstrated by the Covid19 pandemic, there are numerous limitations to using big data. In certain [national contexts](#), the use of big data was quite effective in predicting and limiting the number of new cases, but in many others, it could not capture all social dynamics or environmental conditions that would lead to the spread of the virus.

That said, the use of big data remains popular within many national contexts, considered an effective tool to strengthen the capacity of the state in its delivery of public goods. New ways of policing, and of tracking and predicting crime with Big Data have proliferated in recent years and this comes with immense new challenges because there is hardly such a thing as a ‘purely algorithmic’ reliance on Big Data.

### **Big Data and governance**

In a chapter entitled “Scientific Government” in his 1931 book, [The Scientific Outlook](#), Bertrand Russell argued that the increase in knowledge made governments able to achieve much more with the help of the scientific community. Yet, Russell cautioned against a simplistic appraisal of scientific input to government, and against the ‘wrong-headed and anarchical cranks’. The previous year, the opening article in [Nature magazine on 6 September 1930](#) had contemplated a similar predicament:

“The practical problem of establishing a right relationship between science and politics, between knowledge and power, or more precisely between the scientific worker and the control and the administration of the life of the community, is one of the most difficult confronting democracy.”

The 21<sup>st</sup> century brings additional complications in the context of Big Data. Firstly, the methods behind Big Data are premised on heightened levels of surveillance, and secondly, the providers of the data are not necessarily data scientists in public establishments but most often private companies selling information to state agencies. The case of policing is evocative.

[Big Data technology](#) is allowing the police to become more aggressively proactive. Surveillance technologies allow the police to visualize crime in new ways, as well as to track physical movements, digital communications and ‘suspicious associations’. Over 60 American police departments use some forms of “predictive policing” in daily operations. For example, in Chicago, a “heat list” defined by an algorithm ranks people who are at risk of becoming victims or perpetrators of gun violence. As a result, the police prioritizes particular neighborhoods and districts, or monitors the activity of certain people more closely. Furthermore, with new technologies of surveillance such as body cameras, high-resolution surveillance and automated license-plates readers, the officers themselves see their roles transformed, to become, additionally, data collectors.

Then, there is the complex issue of the quality and reliability of the data collected, as many instances have shown clear cases of disproportionate targeting of certain population groups, and racial injustice

committed in the process. The concern that [‘Big Data blacklisting’](#) may lead to false-positive stops, investigations and arrests is not unfounded, and is an issue that has been discussed in many legal debates, including at the Supreme Court. The lack of transparency in this process is unsettling because government decisions made in the absence of fairness and accountability is a direct affront to the protections of due process.

Adding an extra layer of contention to this issue is the role of private companies which combine digital information from social media and the internet of things with criminal data, which they sell to law enforcement agencies.

This issue was already flagged by the American Civil Liberties Union, taking issue with the Chicago Police Department “heat list”, which associated innocent people with criminal behavior. The large-scale [sharing of data](#) between police departments and the private sector has raised concerns for civil liberties – while simultaneously enabling companies that sell tools for data analysis to make massive profits. For instance, one company, Palantir Technologies, which uses data-sifting techniques to map social networks of extremists, money launderers, or murderers, was allowed by law enforcement agencies in the Salt Lake City area to have access to 40,000 arrest photos, 520,000 case reports and information including airport data – all to build maps of suspected criminals.

Predictive law enforcement is publicly justified to be in the general interest by saving up resources for other activities or for identifying various other problems. However, a cost-benefit analysis risks overlooking the profound ways in which Big Data tools can be detrimental not only to human dignity and civil liberties but also to the meaning of the social contract.

In the next two sections, I provide a neurophilosophical account of the real and likely consequences of Big Data for human dignity and governance, and the inherent risks to the social contract.

### **Impact of Big Data on the Nine Human Dignity Needs**

Big Data is not only poised to erode the relationship between citizens and governments but it is challenging for human nature and human dignity largely speaking. [Human dignity](#) has been considered, inter alia, foundational to the data protection regime in Europe.

With insights from neuroscience, and as detailed in other [posts](#), I previously described human nature as emotional, amoral, and egoistic. We are highly *emotional* beings, with emotionality and emotional-process regions of the brain closely connected with a host of cognitive skills that philosophy previously attributed to ‘reason’, including (moral) decision-making, learning and memory formation. We are also *amoral* in the sense that we are neither innately moral, nor immoral, although we do have some inborn predispositions, such as the predisposition for survival, which pushes us to pursue those actions that likely maximize our chances of survival. The pursuit of survival of the self is a basic form of *egoism* and is a defining and powerful element of our nature.

Understanding the fundamental features of our human nature has important consequences for governance. Because our nature is highly malleable, easily swayed by emotions and by the pursuit of survival, the only sustainable governance models are those that establish the conditions for the best in our nature to thrive, and for sociality to emerge and sustain political life. The most important prerequisite for good governance is human dignity because dignity is fundamental to human existence and in many ways, even more so than other political freedoms and rights. What I mean by dignity is not simply the absence of humiliation, but a more comprehensive set of nine [‘dignity needs’](#), which includes: reason, security, human rights, accountability, transparency, justice, opportunity, innovation, and inclusiveness.

The age of Big Data brings about significant challenges to all of these dignity needs.

*Reason*, which I do not define in neuroscientific or philosophical terms here, but as a reflection of how important dogma is to a society, and the extent to which governments rely on facts as opposed to claiming monopoly on the truth, can be severely affected by Big Data. Big Data comes, as mentioned, with epistemological implications and it is a very powerful tool to create new approaches to governance; yet governments and public institutions may end up committing to the tools and conclusions of collected data while losing sight of the many shortcomings, biases and even randomness that underscores algorithmic data.

*Security* is a fundamental condition for limiting the possibility of fear-induced and pre-emptive violence and must be accounted for in governance because it is linked to our emotionality. Risks of discrimination in data processing, or excessive surveillance exacerbate insecurity and vulnerability across society, or against certain social and ethnic groups, which can be deleterious for social cooperation.

*Human rights* are poised to be strongly affected by Big Data, which could even lead to a pushback on gains and advances achieved in previous decades. Data protection risks being severely compromised and the risks of discrimination or biases in predictive analytics can undermine both individual and political rights.

*Accountability* is critical for fostering trust in institutions because it guarantees the provision of justification for actions and decisions. However, the use of algorithms that operate on Big Data risks weakening the mechanisms for accountability and for *transparency*, a closely related concept crucial for mitigating against forms of discrimination. By their nature, [algorithms](#) are 1. complex both technically and structurally (requiring complex processes of data collection, encoding by a programmer in an algorithmic language before translation into a machine-readable binary sequence etc.), and 2. often operating on a random group level. These two conditions weaken the claim to accountability and transparency, and makes it highly difficult for the general public, which may be affected by the algorithm but does not understand algorithmic language, to understand the algorithm's behavior. In reality, this is difficult even for computer scientists at times, given the possibility of algorithms behaving unpredictably.

A closely related concept here is *justice*, which is critical to human nature and is at the heart of the foundation of the social contract. Big Data leads to serious risks of unfair treatment, profiling and can compromise the judicial process especially when tools such as a [computerized risks assessment](#) are involved, which calculates chances to reoffend and has on occasions led to especially harsh sentences.

*Opportunity* is a dignity need linked to our 'egoism' because it refers to the ability of the state to ensure access to essential resources for self-sustainability and implicitly, for survival. Big Data can curtail opportunity by entrenching stigma and biases in predictive analytics tools, which may '[forecast](#)' [troublemakers](#) for the community. This could, for instance, disqualify certain individuals from employment opportunities, an issue that has been raised before in [countries](#) that rely on algorithmic systems in public and private institutions.

*Innovation* is paramount to our nature and is critical in enabling self-expression. There is a real risk of self-censorship in the age of Big Data and mass surveillance. Artists, writers, innovators may be increasingly pushed to self-impose limits on their online and offline engagements, a trend already noticed following allegations of NSA surveillance. A 2013 [study](#) found that many US writers chose self-censorship, with 28% claiming they curtailed social media activities and 24% reported avoiding

discussing certain topics on the phone or by email. Similar [findings](#) were reported in 2018 for Scotland-based writers.

*Inclusiveness*, which is paramount to any governance system, requires sustained policy efforts to eliminate forms of marginalization which fuel social divisions and resentment. Not only can Big Data reinforce already existing discriminatory practices, it can further silence those who wish to [report](#) on injustices and discrimination, thus making it even more difficult to bring issues to public attention and hope for policy solutions.

These consequences to human dignity and governance will lead to alterations of the meaning of the social contract and the principles upholding it. In the long run, the terms of the social contract may need to be revised.

### **The social contract: neurophilosophical perspectives**

The large-scale collection of data about individuals raises concerns of imbalances of power in society between citizens and tech companies. The [social contract](#) presupposes, at its foundation, a certain *shared vulnerability* and rough *equality of stakes* among the members of society. However, a private entity in possession of large amounts of personal data has ‘predictive powers enabling it to have an unequal stake in the shared system’. The possible errors in overreliance on Big Data call for a new social contract, which ensures more accountability and that all predictive models do not reinforce stereotypes.

Social contract theories have a long history and date back to the earliest philosophical approaches to the state and political authority. This includes ancient Egyptians, Greek, Roman, Chinese and Indian traditions, and in more recent history the writings of Hobbes, Locke, and Rousseau. In various ways, these writings explored the limits of freedom and the legitimacy of the sovereign’s powers.

[Socrates](#), depicted in the Platonic dialogue, *Crito*, explained the compelling reason why he must stay in prison and accept the death penalty – rather than escape and move to another Greek city – and that was because of an overwhelming obligation to obey the Laws because the laws had made his entire existence possible thus far. For Hobbes, the social contract ensured a way to limit the effects of the inherent selfishness in each individual in the state of nature, which would lead to terrible outcomes were it not for the Leviathan. The social contract was the only way to avoid the state of nature, which was no way to live, according to Hobbes. For Rousseau, the social contract was also tied to democracy and endowing people with rights and freedoms.

The common thread in ideas of the social contract is that human beings ultimately choose to give up some freedoms in order to enjoy life in a political order that exists for the interest of *all* citizens and where life and property are protected. Underlying this contract is therefore the belief that it is a fair and [just](#) exchange that benefits everyone.

The inherent imbalances in power created by Big Data predictive tools pose a fundamental challenge to the sense of fairness underscoring the social contract, and thus undermine its acceptability. This sense of fairness is critical to human nature in profound ways.

As [emotional, amoral and egoistic](#) beings, highly influenced by circumstances and geared towards survival (and actions that maximize our chances of survival), concerns about justice and fairness are highly important to us, and often even linked to survival. The [sense of fairness](#) – and what we perceive as fair – and of justice, as well as understandings of harm reduction and cooperative behaviors are regulated neurochemically by increases or depletions of serotonin. Additionally, inequality aversion is part of the larger picture of interconnected responses to fairness. [More recent research](#) also

suggests that humans are keen to pursue actions with personal costs in order to punish those who violate social norms.

Going forward, and as citizens will keep on feeding into the data collected and analyzed by companies and states, a stronger sense of fairness must be reintroduced in the fabric of the social contract. Indeed, the social contract can only remain an enduring bond if it is accompanied by constitutional guarantees of enhanced data protection and transparency. Vitiating by a growing sense of deceit and unfairness, the social contract will only become weaker and no amount of data can restore its acceptability.



### **Nayef Al-Rodhan**

*[Prof. Nayef Al-Rodhan \(@SustainHistory\)](#) is a Neuroscientist, Philosopher and Geostrategist. He is an [Honorary Fellow at St Antony's College, University of Oxford](#), and Senior Fellow and Head of the Geopolitics and Global Futures Programme at the [Geneva Centre for Security Policy](#), Geneva, Switzerland. Through many innovative books and articles, he has made significant conceptual contributions to the application of the field of neurophilosophy to human nature, history, contemporary geopolitics, international relations, cultural studies, future studies, and war and peace.*