12

SOFTWARE, ALGORITHMS AND DATA

Key questions

- What is the contribution of 'software studies' to more conventional social and cultural perspectives on digital society?
- How can we tease out the sociologically interesting aspects of inherently digital things such as links, likes, and search engine results?
- What social and economic role do algorithms play?
- How can the phenomenon of 'big data' be approached from a critical sociological perspective?

Key concepts

Software * algorithms * the link economy * the like economy * calculated publics * big data

Software is a word most commonly used to refer to computer programs and applications. Software is the set of instructions, the code, which instructs technological objects to function in the desired way, as opposed to hardware, which is the physical technological objects in the form of computers, mobile phones, televisions, or refrigerators. The first published use of the term software in relation to computing was in an article from 1958 by statistician John Tukey (1958: 2):

> Today the 'software' comprising the carefully planned interpretive routines, compilers, and other aspects of automative programming are at least as important to the modern electronic calculator as its 'hardware' of tubes, transistors, wires, tapes and the like.

His point, then, was that not everyone who uses hardware gadgets has to learn all the details of what goes on 'under the hood', and of course, this is still the case. When we use a computer, mobile phone, or a social media application, we generally do not think very much about the logic that governs these things. We just use the things without reflecting on the details of how they were programmed or why. Even though those details and choices may have a huge effect on what we can do with the tools, and how we do it, it becomes impractical to consider their underlying logic.

In recent years, however, researchers have become more and more interested in looking at *software* in a more general sense, as software has become a force that structures and enables much of our contemporary world. New media theorist Lev Manovich (2013: 6) writes that software is 'the engine of contemporary societies'. Friedrich Kittler, a media theorist and literary scholar, said in an interview (Griffin et al. 1996: 240) that software had become increasingly important for understanding culture. He said:

> I can't imagine that students today would learn only to read and write using the twenty-six letters of the alphabet. They should at least know some arithmetic, the integral function, the sine function — everything about signs and functions. They should also know at least two software languages. Then they'll be able to say something about what culture is at the moment.

In a vein similar to that of Aakhus and Katz's theory of apparatgeist, discussed in Chapter 11, Kittler underlined the importance of critical analysis of the 'essence' of computers as a complex phenomenon. First, he wrote that 'software does not exist as a machine-independent faculty'. This means that software can't be studied in separation from hardware. Second, he claimed 'there would be no software if computer systems were not surrounded any longer by an environment of everyday languages'. This means that software is not strictly confined to computers. Its logic and effects bleed out into the rest of society — and vice versa.

In this chapter, I discuss the emergence of the research field of *software studies*. Until recently, the social sciences and the humanities have largely ignored the phenomenon of software — the underlying code of digital society — and instead focused more broadly on the social and cultural effects of digital media, as discussed elsewhere in this book. Recently, however, there has been an increased interest in the critical analysis of different aspects of how software enables and limits various social practices, and how software, defined as a general system of signs and functions, is shaped by, and shaping, social interaction. This covers a broad variety of objects of study, spanning how code, files, copies, visualisations, functions, glitches, interfaces, bugs, and so on 'leak out of the domain of logic and into everyday life' (Fuller 2008b: 1). This chapter deals with the important, but largely invisible, role played by *algorithms* in digital society, and with concepts such as *calculated publics*, *the like economy*, and *big data*.

A UNIVERSAL ENGINE ON WHICH THE WORLD RUNS

Software as an object of study is a moving target. Rapid technological development, and the acceleration of consumer capitalism has meant that, as Manovich (2013: 2) writes, 'the world is now used to running on web applications and services that have never been officially completed but remain forever in Beta stage'. A substantial part of all of the applications and services that people interact with in their everyday lives run on remote servers, meaning that they can be invisibly — or secretly — updated anytime. This is often the case, as services that aspire to become the operating system of digital society — such as Google and Facebook — update their code on a daily basis. As Manovich (2013: 2–3) puts it: 'Welcome to the world of permanent change — the world that is now defined not by heavy industrial machines that change infrequently, but by software that is always in flux.'

As cultural researcher Matthew Fuller (2008b) explains, an important task of software studies is to show that software is a vital object of study, as well as an area of practice, for researchers and thinkers in fields that one would not conventionally associate with 'software', in the narrow sense. While disciplines such as computer science, informatics, and related fields that work on the interface between computer and human have done lots of important work on the topic of software, it has long not been researched at all in cultural and social studies. Manovich (2013: 2) makes a convincing case, arguing that no matter what social and cultural things we do with digital devices — play, watch, listen, write, blog, tweet, call, talk, email, edit, take photos, film, and so on — we are all the time using software, and:

> Software has become our interface to the world, to others, to our memory and our imagination — a universal language through which the world speaks, and a universal engine on which the world runs. What electricity and the combustion engine were to the early twentieth century, software is to the early twenty-first century.

Software tends to become a transparent or invisible aspect of digital society, in spite of its crucial role for its functioning. In the same way that hegemonic power and ideologies are naturalised, uncriticised, and spontaneously consented to, software also has an 'ideological layer' (Fuller 2008b: 3). Even though software is often extremely useful and even empowering, scholars in software studies remind us that much software — as a by-product — also defines social relations in certain ways that become systematic and impossible to alter once they are set. For example, consider the worries, discussed in Chapter 3, of Jaron Lanier that abstract automated functions will remove humanity. The 'user-friendly' software of social networking will kill off the personal, varied, often nicely strange homepages from the days when 'the web had flavor' (Lanier 2010: 15). Software, in this case, leads to a form of self-reduction:

> The binary character at the core of software engineering tends to reappear at higher levels. It is far easier to tell a program to run or not to run, for instance, than it is to tell it to sort-of run. In the same way, it is easier to set up a rigid representation of human relationships on digital networks: on a typical social networking site, either you are designated to be in a couple or you are single (or you are in one of a few other predetermined states of being) — and that reduction of life is what gets broadcast between friends all the time. What is communicated between people eventually becomes their truth. Relationships take on the troubles of software engineering. (Lanier 2010: 71)

Examples like this illustrate how software, something that is often defined as immaterial, actually has very material consequences in digital society. The design of software operates on many levels. It defines the characteristics of languages and interfaces; it enables certain kinds of use, and disables others.

EXERCISE

Look with this new perspective at a website or app that you regularly use in your everyday life. Try to make it the subject of a rudimentary software analysis. Distance yourself from the position you are in now where the functions and resources that make up the site or app are very familiar — maybe nearly transparent — to you. Instead, pose critical questions about it: What is it actually designed to achieve? Which functions are offered, and how? What other things *could* it have been designed to achieve? Which other functions *could* have been included? What does it seem to assume about, or expect from, its users in terms of interests, goals, skill level, gender (or any other dimension that will bring out interesting insights)? Which users are excluded? Is the site or app based on any particular values or convictions? Thinking in this way will make the software come into view so that it can be critically analysed.

FROM HITS TO LINKS

Software studies are interested in natively digital things. These are things that would not exist if it was not for digital media and the internet. So, while for example a conversation between two people, or a television broadcast, can happen either through digital tools and platforms or without them, there are some phenomena that are distinctive to the digital. New media researcher Richard Rogers (2013: 25) argues for the importance of 'following the medium' — more about that in Chapter 16 — and looking closer at what is specific to the digital. Beyond analysing online culture, or what happens to society as it becomes digital, there is also a need for research to capture and analyse natively digital things such as 'hyperlinks, tags, search engine results, archived websites, social networking sites' profiles, Wikipedia edits', and so on (2013: 19). Such an approach is about an analysis of what communications researcher Tarleton Gillespie (2010) names as 'the politics of platforms'.

New media researchers Carolin Gerlitz and Anne Helmond (2013) have explored how different types of 'web-native objects' have organised value production — economic and other — online. In the mid-1990s, in the days of web 1.0 (see Chapter 2), the most important objects were the hit and the hyperlink. During this period, the number of *hits* on a website became widely used as the standard metric to measure user engagement and website traffic. Many websites had 'hit counters' that displayed how many visitors a page had attracted, based on the number of computerised requests to see the page. In the late 1990s, however, this standard was replaced as Google, which was then a new type of search engine, introduced a new way of measuring impact by combining hits and *links*. This was a watershed in the history of the internet as it gave rise to a new web economy with search engine rankings at the centre.

PageRank

Famously, Google introduced the analysis algorithm of PageRank, which was developed in 1996 by founders Larry Page and Sergey Brin as part of a research project at Stanford University. PageRank — the name playing off both the name of Page and the notion of a web page — calculates the relative importance of a page according to a rather intuitive logic:

> a page can have a high PageRank if there are many pages that point to it, or if there are some pages that point to it and have a high PageRank. [...] PageRank handles

(Continued)

(Continued)

> both these cases and everything in between by recursively propagating weights through the link structure of the Web. (Brin & Page 2012: 110)

The basic idea, in other words, is that a page A can have a higher PageRank than a page B, even if B has more links pointing to it. This is because A may be linked fewer times but by more important pages. The point with this was that by using PageRank in addition to conventional text indexing, one would be able to generate much more accurate search results. Google's algorithm brought in a focus on the relational value of sites and thereby shifted the way in which the value of web resources was determined away from the hit and towards the link as the main measure of relevance. This was done according to the logic of PageRank where links have different value depending on the authority of the source.

This made links into a commodity in a new form of web economy where search engine optimisation (SEO) emerged as a key practice for actors who wanted to capture people's attention online. SEO involves practices such as the careful choice of keywords for the site's meta description, the creation of content that includes frequently searched words, frequent updates to lure the automated crawlers of the search engines to re-index the site, and so on. Many emerging SEO practices went beyond merely helping the search engines build appropriate indexes, instead bordering on spam — so-called spamdexing. One such practice to deliberately manipulate the indexing process is carried out through so-called 'link farms' — a group of websites that all link to each other in order to boost their PageRank. Links have a direct value in digital society, and they can therefore be seen as a 'pseudomonetary unit' (Rettberg 2005: 526). And it's not only the links themselves that have value, but the knowledge about the relationships between content that became a 'prime real estate' (2005: 525). Links were increasingly exchanged in strategically reciprocal ways. Jill Walker Rettberg (2005: 526) explains how the link economy functions:

> When I link to B, I give B a link. That link translates into a precise (though undisclosed) value in Google's PageRank and in other indexing systems […]. The link has a clearer value to B than the content of B's page has to me or to my readers. I pay B for B's content with my link. This instrumental view of links

> does not exclude its other qualities. Many people creating or following links on the Web link generously, carefully, or haphazardly but without thinking of the economy of links and their value.

As a consequence of the link economy, link bartering, loosely organised systems of linking someone and being linked in return, was made more formal through phenomena and functions such as webrings and blogrolls. Such practices subverted Google's 'objective' measurement of links, and when they got too overtly strategic, they were sometimes labelled as 'link slutting' or 'link incest' (Rettberg 2005: 528). It was frowned upon to shamelessly or inappropriately sell your integrity for links. But gradually, there was also an increasingly open exchange of links for real-world money. A black market for links emerged, where people could pay to be linked by link farms, circles, and other technological agents designed to do nothing but link to others. Consequently, Google developed different practices to police and ban such activities. Of course, it was in Google's interest to protect the integrity of its system, since the map of the Web that they were, and are still, developing is priceless, not only for the generation of as 'good' search results as possible, but also for the ability to personalise searches and — by extension — ads (see Chapter 9). As more and more of our online activities are tied into our user profiles with corporations like Google, Facebook, or Apple, these actors will have more and more data about us in their rapidly-expanding databases.

THE LIKE ECONOMY

After the arrival of the social web and, consequently, social media, there were further changes to the attribution of value to sites and content. Initially, the participatory features of web 2.0 made it possible for users who were increasingly engaged in the creation of their own content to be more active also in the creation of connections between sites, accounts, and platforms. The early renditions of the link economy had been predominantly based on expert recommendations and aggregation engines such as Technorati and Blogpulse. However, as Richard Rogers (2005: 27) explains, 'the blogsphere became a new kind of collective, aggregated source — one freed from the "tyranny of (old media) editors"'.

The emergence of 'social buttons' that could be placed on any website were a further development towards more participatory linking practices. These buttons enabled the submission of, or voting for, posts on platforms such as Digg and Reddit, which introduced sharing buttons in 2006 (Gerlitz & Helmond 2013: 1351). Many other platforms followed suit and offered different social buttons which allowed for a variety of predefined user activities: bookmarking, voting, recommending, sharing, and tweeting, and counters which showed how many times they had been clicked.

Buttons

Digital aesthetics scholar Søren Pold explains that buttons in web interfaces and apps have a certain social power, since buttons 'signify a potential for interaction' and because buttons feel very real and definite. Pold (2008: 32) writes:

> There is an analog connection between pressing the button and, by the force of one's finger transmitted through a lever, changing the state of the apparatus — as in old tape recorders, where one actually pushed the tape head into place with the button. The computer interface does away with the analog mechanical functionality, but the function of buttons here is to signify the same stable denotation, even though its material basis is gone. That is, interface buttons disguise the symbolic arbitrariness of the digital mediation as something solid and mechanical in order to make it appear as if the functionality were hardwired.

The major transformation came with Facebook's introduction of the like button in 2009. The now classic thumbs-up button was created in order to be a shortcut for comments, and to replace short affective statements such as 'Congrats!' or 'Awesome!'. Since 2009, there has been a longstanding debate on the absence of a 'dislike' button. Critics have argued that a button for positive sentiment only, works to support commercial interests, such as building brands or promoting products and services. Mark Zuckerberg, head of Facebook, said in 2014:

> Some people have asked for a dislike button because they want to say, 'That thing isn't good.' And that's not something that we think is good for the world. So we're not going to build that.[1]

However, in February of 2016, Facebook introduced a wider range of 'reaction' options: Like, Love, Haha, Wow, Sad, or Angry.

The social act of liking — or otherwise 'reacting' — can be performed on most things on Facebook, through actions such as status updates, shared photos, shared links, or comments. As with the social buttons that preceded it, from the very beginning, the like button had a counter, and also listed the names of those who had clicked it.

[1] www.slate.com/articles/technology/future_tense/2014/12/facebook_dislike_button_why_mark_zuckerberg_won_t_allow_it.html.

A year later, in 2010, Facebook launched an external like button that could be used as a plugin by any site owner, 'potentially rendering all web content likeable' (Gerlitz & Helmond 2013: 1352).

This innovation made links — the main currency of the link economy — less interesting and instead put the focus on how 'liking', or performing other preset 'reactions', transforms user interactions into comparable and actionable forms of data. The emerging like economy facilitated a more social web experience, where being liked and seeing what others like enables new forms of engagement. But, Gerlitz and Helmond (2013) argue, it also creates 'an alternative fabric of the web in the back end'. In this obscured dimension, specific relationships are created 'between the social, the traceable and the marketable'. So, while the link economy bore traces of democratisation, as it was a system where anyone could link to anyone else, the like economy means a recentralisation. Many people are involved in the 'liking' part of the like economy, but most of them lack full access to the data they are part of producing. Instead of the patterns generated through mutual linking practices, the like economy presents an alternative fabric of the web, which is organised through data flows that emanate from social media platforms such as Facebook.

The like button, embedded both inside and outside Facebook, is an example of a 'tracking device', which establishes new markers of relationships online that go beyond the conventional hyperlink between websites. So, the fabric of the like economy is not organised through relationships between websites, but instead through third-party tracking devices, linked to data mining services. Fundamentally, the digital artefact of the Facebook like, or reaction, button sets up a particular relationship between the social and the economic dimensions of society. The widespread use of Facebook, the prominence of the like button, and appearance of the embedded like button throughout the internet makes it possible to gather large amounts of valuable user data.

EXERCISE

Try to reflect upon what a 'like' is to you. From the perspective described above, the like button is a tracking device for generating economic value. From another perspective, it can be a shorthand for conveying a positive sentiment. One could also imagine that the meaning of a like is very contextualised. In the cases when you 'like' something, do you ever think about how that click is going to be interpreted by others? Is 'liking' just something that we do, compulsively? Is the meaning of the like taken for granted? May the like even be an empty signifier, in the sense that its meaning is not fixed?

ALGORITHMS

Algorithms play a key role in the softwarisation of society. From a strictly computational point of view, algorithms are mathematical procedures that are performed in a controlled fashion on data in order to be able to present an output in the shape of other forms of data. Algorithms are the important procedural logics that undergird all computation. The storage and reading of data, the application of procedures to it, and the delivery of some form of output can be done by hand as well, but the way in which digital society relies on computational tools has turned automatisation and digital routines into a social key mechanism, which governs the flows of information we depend on. Media and communications scholar Taina Bucher (2012: 1), writing about 'programmed sociality', is among those who have shown that algorithms can 'establish certain forms of sociality' by way of how they 'produce the conditions for the sensible and intelligible'. We rely on search engines for the navigation of massive informational databases, or the entire web, and in this process, algorithms help us decide and select what information is important to us. For example, many online services and platforms have recommendation algorithms that suggest to us which book to buy, which Twitter users to follow, what TV series to watch next, who to 'friend', which content is 'hot' or 'trending', and so on. In doing their work, algorithms highlight some bits of the world, while hiding others. For Gillespie (2014: 168), the role of algorithms in society is important since where we may have previously relied on credentialled experts, scientists, 'common sense', or religious authority for correct knowledge about reality, we have now turned to algorithms.

Obviously, not everyone welcomes such 'behind the scenes' mechanics. Algorithms may be beneficial but they may also be exploited to manipulate users. In February 2016, Twitter announced the launch of its 'algorithmic timeline', which was followed by a storm of protests from its users. The change meant that the service would depart from the presentation of tweets in reverse chronological order, in favour of the provision of algorithmically produced tweets, based on user activities. *Wired* magazine contributor Brian Barrett presented an analysis of the changes that indicated that new and uninitiated users might be aided by the new, more accessible method of presenting tweets — 'isolating the signal from the noise'.[2] However, as Barrett wrote, power users who were comfortable with the platform, and had a longer history and familiarity with the original reverse chronological presentation became suspicious and launched hashtags such as #RIPTwitter. So, while some individuals, in some contexts, might be perfectly happy to have their content feeds 'refined' by algorithms, other individuals in other contexts may feel that the very same algorithms 'destroy' their feeds.

[2] www.wired.com/2016/02/a-twitter-algorithm-wont-ruin-anything/

As discussed previously, it is important to carry out critical social analyses of software and algorithms, because they have a certain unquestionable quality to them. Even if we know that an algorithm selectively puts our YouTube start page together, it is still somewhat natural to perceive it as being *the* YouTube start page. But, as Gillespie (2014: 169) argues, algorithms are socially constructed, rather than objective and precise:

> A sociological analysis must not conceive of algorithms as abstract, technical achievements, but must unpack the warm human and institutional choices that lie behind these cold mechanisms. I suspect that a more fruitful approach will turn as much to the sociology of knowledge as to the sociology of technology. [...] This might help reveal that the seemingly solid algorithm is in fact a fragile accomplishment.

Furthermore, Gillespie writes, the algorithms that underpin digital society, the internet, and social media platforms all contribute to the production and legitimisation of knowledge, according to a logic based on assumptions that are very specific. This is why it is important to examine algorithms as a key feature of the media ecosystem of digital society. What are these specific assumptions in relation to given algorithms and contexts, and what are their social and political ramifications?

Googlisation

Philosopher Michel Foucault wrote that knowledge is closely related to power. He said that the knowledge of the world that is established 'tends to exercise a sort of pressure, a power of constraint upon other forms of discourse' (Foucault 1972: 219). As Rettberg argues, this is also highly pertinent to the political economy of links. Links may be useful, functional, or provide us with happiness, for example, but links are also part of a power structure, which must not be ignored. Links define what can be found and so they define knowledge, knowledge, which, once again, is power. Cultural historian Siva Vaidhyanathan (2011) thinks that there has been a *googlisation* of everything, and that in hindsight it might have been a better idea not to put the entire 'human knowledge project' in the hands of a single corporation. We must not assume, he argues, that Google will deliver to us what we 'actually need'. Even though Google might have grandly promised not to 'be evil', it is still big business. Vaidhyanathan argues that:

(Continued)

(Continued)

> About the same time that Google started, we could have coordinated a grand global project, funded by a group of concerned governments and facilitated by the best national libraries, to plan and execute a fifty-year project to connect everybody to everything. (2011: 203)

Activist and author Eli Pariser is also worried about the future. In *The Filter Bubble* (2011), he writes that the evolution of Google and social media, with their underlying algorithms, has ushered people into a personalised and filtered world, where all search results and other information that they are served reinforces their pre-existing values as well as their view on the world. This compartmentalisation and customisation erodes the common ground that people need to share in order to build community and to engage in democratic politics. The googlisation of social reality brings many problems; our beliefs are seldom challenged, which reduces our drive and desire to try to understand others and to incorporate alternative ways of thinking and seeing the world. An element of randomness is needed if we are to be open to discovery.

EXERCISE

Try to break out of the 'filter bubble' by experimenting with different search queries in different search engines with different settings. Choose a search query and enter it into the search field at google.com. Take note of the top search results. Enter the same query at — for example — google.jp, google.ru, google.tn and google.co.uk. Take note of the respective search results. Try the same query at bing.com, duckduckgo.com, yandex.ru, or others. If you like, you can play around with settings for the different engines as well. Note your top search results throughout. When you have finished, analyse the differences and overlaps in the search results. What conclusions can you draw from this?

CALCULATED PUBLICS

A consideration of exactly what algorithms might include or exclude is a vital area for research. In practice, algorithms and the databases upon which they are applied are seen as one and the same phenomenon. But, as Gillespie argues, from an analytic point of view, the two must be studied separately. Algorithms are meaningless without data, and before an algorithm can generate any type of output or result, some information must be collected as input. This process always includes a set of choices about what should be collected and how it should be ordered and 'readied for the algorithm'. The collected data must always be cleaned and ordered into some form of matrix or other readable structure. Furthermore, data — even before they are collected — can be trimmed, primed, and vetted by owners of sites and platforms. Content which is deemed to be 'problematic' can be removed altogether, but it can also be algorithmically demoted in subtler ways. YouTube, for example, withholds 'suggestive content' from lists of most watched videos, or in other recommendation systems. Generally, there is a process of tidying up data. Gillespie (2014: 172) says that:

> Indexes are culled of spam and viruses, patrolled for copyright infringement and pornography, and scrubbed of the obscene, the objectionable, or the politically contentious.

Such tidying is of course necessary, and even helpful, to a certain degree. But it is still valuable to reflect upon this as a form of subtle censorship and to analyse what the choices mean — especially when algorithms have an aura of automation and objectivity. It is also important to consider the social character of algorithms – after all, someone designed and devised them. Rather than thinking only about the effects of algorithms, it might be more fruitful to scrutinise them in terms of their entanglement with the lived world of their creators and users. The entanglement of algorithms with users leads to the rise of what Gillespie (2014: 188–189) calls 'calculated publics'. He explains how algorithms create types of publics that don't really exist in the conventional sense:

> When Amazon recommends a book that 'customers like you' bought, it is invoking and claiming to know a public with which we are invited to feel an affinity — though the population on which it bases these recommendations is not transparent, and is certainly not coterminous with its entire customer base. When Facebook offers a privacy setting that a user's information be seen by 'friends, and friends of friends,' it transforms a discrete set of users into an audience — it is a group that did not exist until that moment, and only Facebook knows its precise membership. These algorithmically generated groups may overlap with, be an inexact approximation of, or have nothing whatsoever to do with the publics that the user sought out.

Similarly, Twitter's algorithm which shows live 'trending' topics within a certain national or regional public also leads to the definition of a highly constructed public, shaped by criteria that are specific and unspecified at the same time. Gillespie defines the notion of *calculated publics* in relation to that of networked publics (see Chapter 2). His main point is that there is a friction in digital society between the — networked — publics that are forged by users through their social interaction with each other and the calculated, somewhat artificial, publics that are generated through algorithms.

Digital sociologist Deborah Lupton (2016) writes that the move towards tracking and monitoring users' movements within and across digitally networked tools and platforms has given rise to new ways of conceptualising people and what they do. Instead of conventional socially and culturally embedded identities, we develop 'data selves' that are configured by the bits of information we generate and collect. As Lupton suggests, one might argue that rather than having a traditional sense of selfhood, people today have started to understand themselves as an assemblage of data. As we are becoming data, we must increasingly understand ourselves as such. This process is further extended with the development of our 'quantified selves' as a consequence of the increased use of techniques of 'lifelogging', personal informatics, and personal analytics, through apps, and wearable technologies such as smart watches and wristbands.

CHALLENGING BIG DATA

Big data has been defined in several different ways since the term was first used in the mid-1990s to refer to the handling and analysis of massively large datasets. According to a popular definition, big data conforms with three Vs: it has volume (enormous quantities of data), velocity (is generated in real-time), and variety (can be structured, semi-structured, or unstructured). To this, various writers and researchers have suggested a number of other criteria be added, such as exhaustivity, relationality, veracity, and value. During a review of a number of big data sets in order to find their common traits, geocomputational researchers Rob Kitchin and Gavin McArdle (2016) found that the two most important characteristics of big data are velocity and exhaustivity. This means that big data captures entire systems rather than samples (exhaustivity) and that it does so in real-time (velocity). Crawford and boyd (2012) think that 'big data' is in fact a poorly chosen term. This is because its alleged power is not mainly about its size, but about its capacity to compare, connect, aggregate, and cross-reference many different types of datasets (that also happen to be big). They define big data as:

> a cultural, technological, and scholarly phenomenon that rests on the interplay of: (1) Technology: maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets. (2) Analysis: drawing on large data sets to identify patterns in order to make economic,

> social, technical, and legal claims. (3) Mythology: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy. (Crawford & boyd 2012: 664)

From a critically sociological perspective, Lupton (2014: 101) argues that the hype that surrounds the new technological possibilities afforded by big data analyses contribute to the belief that such data are 'raw materials' for information — that they contain the untarnished truth about society and sociality. In reality, each step of the process in the generation of big data relies on a number of human decisions relating to selection, judgement, interpretation, and action. Therefore, the data that we will have at hand are always configured via beliefs, values, and choices that '"cook" the data from the very beginning so that they are never in a "raw" state'. So, there is no such thing as raw data, even though the orderliness of neatly harvested and stored big data sets can create a mirage to the contrary.

Sociologist David Beer (2016: 149) argues that we now live in 'a culture that is shaped and populated with numbers', where trust and interest in anything that cannot be quantified diminishes. As Crawford and boyd (2012: 665) argue, the mirage and mythology of big data demand that a number of critical questions are raised with regards to 'what all this data means, who gets access to what data, how data analysis is deployed, and to what ends'. There is a risk that the lure of big data will sideline other forms of analysis, and that other alternative methods with which to analyse the choices, expressions, and strategies of people are pushed aside by the sheer volume of numbers. 'Bigger data are not always better data', they write, and they use the example of Twitter analysis to demonstrate that the bigness of tweet data does not mean that an analysis of it will necessarily lead to insights about society that are more true than other data and methods:

> Twitter does not represent 'all people', and it is an error to assume 'people' and 'Twitter users' are synonymous: they are a very particular sub-set. [...] For example, a researcher may seek to understand the topical frequency of tweets, yet if Twitter removes all tweets that contain problematic words or content – such as references to pornography or spam – from the stream, the topical frequency would be inaccurate. Regardless of the number of tweets, it is not a representative sample as the data is skewed from the beginning. (Crawford & boyd 2012: 669)

In sum, Crawford and boyd, who see the emergence of big data as part of a more wide-ranging 'computational turn' in culture and society, underline the importance of recognising the rhetoric surrounding big data. We must remember that the design and interpretation of big data is socially constructed, that there is still value to be found in 'small data', and that there are a number of unresolved and problematic ethical issues that surround the use of big data.

FURTHER READING

Manovich, Lev (2013). *Software Takes Command*. London: Bloomsbury.
Manovich presciently called for 'software studies' in his 2001 book *The Language of New Media*. In this volume from 2013, he presents a further development of that idea. Focusing especially on 'media software' (such as Photoshop, After Effects, and Google Earth), Manovich discusses where such software comes from (historically), and how it shapes how media is created, viewed, and remixed.

Fuller, Matthew (Ed.) (2008a). *Software Studies: A Lexicon*. Cambridge, MA: MIT Press.
Software studies often focuses on the influence of software, but this edited volume is also interested in the very material of software. Writers from a wide variety of fields have contributed short texts about key topics such as 'algorithm', 'code', 'copy', 'glitch', and 'pixel'.

Lupton, Deborah (2016). *The Quantified Self*. Cambridge: Polity Press.
Lupton examines the emerging field of self-tracking through digital devices and software. She deals with a set of related issues from a social and cultural perspective. Lupton specifically pays attention to how the large amounts of data generated and collected via self-tracking tend to be collected and used for different purposes by businesses, governments, and researchers.