The limits of computation: A philosophical critique of contemporary Big Data research

Törnberg, P.; Törnberg, A.

Link to publication

Original Research Article

The limits of computation: A philosophical critique of contemporary Big Data research

Petter Törnberg[1] [ORCID] and Anton Törnberg[2]

Abstract

This paper reviews the contemporary discussion on the epistemological and ontological effects of Big Data within social science, observing an increased focus on relationality and complexity, and a tendency to naturalize social phenomena. The epistemic limits of this emerging computational paradigm are outlined through a comparison with the discussions in the early days of digitalization, when digital technology was primarily seen through the lens of dematerialization, and as part of the larger processes of "postmodernity". Since then, the online landscape has become increasingly centralized, and the "liquidity" of dematerialized technology has come to empower online platforms in shaping the conditions for human behavior. This contrast between the contemporary epistemological currents and the previous philosophical discussions brings to the fore contradictions within the study of digital social life: While qualitative change has become increasingly dominant, the focus has gone towards quantitative methods; while the platforms have become empowered to shape social behavior, the focus has gone from social context to naturalizing social patterns; while meaning is increasingly contested and fragmented, the role of hermeneutics has diminished; while platforms have become power hubs pursuing their interests through sophisticated data manipulation, the data they provide is increasingly trusted to hold the keys to understanding social life. These contradictions, we argue, are partially the result of a lack of philosophical discussion on the nature of social reality in the digital era; only from a firm metatheoretical perspective can we avoid forgetting the reality of the system under study as we are affected by the powerful social life of Big Data.

Keywords

Big Data, data analytics, end of theory, computational social science, data-driven science, epistemology

Introduction

The term "Big Data" is used to describe the volume of information produced through the use of technologies like mobile devices, positioning systems, and online services—"[i]n a digitized world, consumers going about their day—communicating, browsing, buying, sharing, searching create their own enormous trails of data" (Manyika et al., 2011: 1). The increasing use of digital services has given social scientists unprecedented access to previously unimaginable data; traces of the lives, dreams, and feelings of hundreds of millions of people. This seems to bring great promises for social scientific work, as the "data deluge . . . is leading us to an ever greater understanding of life on Earth and the Universe beyond . . . [it may] transform the process of scientific discovery. The more data there is the more discoveries can be made" (Rosling, 2010). Some have pointed to a "fourth paradigm" for science, as new algorithmic, computational, and analytical tools produce "gold" from this data resource (Bell et al., 2009; Hey et al., 2009).

[1]Department of Sociology, University of Amsterdam, Netherlands
[2]Department of Sociology and Work Science, University of Gothenburg, Sweden

Corresponding author:
Petter Törnberg, Department of Sociology, University of Amsterdam, Postbus 15508, 1001 NA, Amsterdam, Netherlands.
Email: k.p.tornberg@uva.nl

But Big Data is also associated to a number of scientific challenges, that have led to the competing view that "the glittering promise of online data abundance too often proves to be fool's gold" (Karpf, 2012: 652). Big Data has brought a development that is simultaneously enticing and vexing: a veritable "siren-song of abundant data" (Karpf, 2012) that is causing researchers to flock to the study of phenomena identifiable in growing Big Data, and ignore phenomena not so inscribed. They are met by a host of difficulties, some of which are a natural part of new technologies to which we have yet to grow accustomed, but there are also issues that run deeper than the mere settling of dust (e.g. Boyd and Crawford, 2012: 20). Big Data is not only different in quantity, but also in quality, and it seems that the new shapes of data do not always fit into the holes of old theory. This has resulted in it bringing fundamental issues to the fore: the questioning of epistemological assumptions, discussions of the validity of disciplinary divides, critique of methodological monism, and the rejection of long-trusted simplifications (Kitchin, 2014). Throughout the sciences, similar tensions can be seen as the data deluge causes long submerged epistemological questions to float to the surface.

While the traditional variable-based approaches to social science have struggled with the new forms of data, approaches with their roots in the natural sciences have stepped forth to meet the tide. A central player in the multitude of new approaches is Computational Social Science, located at the "intersection of the social and computational sciences, an intersection that includes analysis of web-scale observational data, virtual lab-style experiments, and computational modeling" Watts, 2013:1. This natural scientific approach has brought with it a range of methods to allow dealing with the complexity of mass-interaction (Conte et al., 2013). The renegotiation of demarcations between the natural and social sciences following from this development seems to in part be leading to a renewed naturalism, sometimes referred to as the "end of theory" (Anderson, 2008) and exemplified by Lev Manovich's (2016) view that "Digital is what gave culture the scale of physics, chemistry or neuroscience. Now we have enough data and fast enough computers to actually study the 'physics' of culture". In this new naturalism, society is subsumed not under the traditional Cartesian-Newtonian paradigm, but instead under the metatheory of particles and flows, and analogues such as "avalanches and granular flows, flocks of birds and fish, networks of interaction in neurology, cell biology and technology" (Ball, 2012: ix).

This situation paints a picture in which Big Data has taken parts of social science, in particular fields like Computational Social Science, towards a new computationally based perspective by underlining the limitations of traditional quantitative methods. As we will argue, their response implies a particular social ontology, which focuses on relations and sees social structures as patterns emerging from underlying local interaction.

This paper looks at what this approach leaves out. By weaving together a series of theoretical views, the paper outlines the epistemic limits of the emerging computational paradigm. We bring back into the contemporary discussion the views that dominated the social sciences in the early days of the digital era—a time when digitalization was seen not as something that would make society more amenable to formal methods, but rather as the precise opposite: as part and parcel of the processes of postmodernity. It was seen as part of a development towards increasing *openness* to society as a system, thereby limiting the usefulness of quantitative approaches. We argue that these theoretical points still hold and should be taken into account within the current relational perspective, but that they also need to be brought up to date with more recent developments of digital technology, focusing on power and the role of digital platforms.

We begin by looking at the impact that Big Data has had on contemporary social science.

Contemporary digital research: Computation and relations

Despite its name, size is arguably not the most defining feature of "Big Data" (e.g. Boyd and Crawford, 2012): the concept rather describes a set of parallel developments in various disciplines, whose common denominator is an increasing proliferation of data sets that have proven difficult to fit into existing paradigms. In the computer industry—first to feel the effects of this development—quantity was indeed among the primary issues, since traditional tools, such as relational databases, proved incapable to deal with new demands emerging from large-scale systems (Manovich, 2011). But in the social sciences, the emerging problems associated to Big Data were different: as Boyd and Crawford (2012) observe, some of the data sets understood as examples of "Big Data" (e.g. some Twitter studies) are significantly smaller than sets understood as "traditional" data (e.g. census data), pointing to the fact that the data quantities in themselves are not the issue—even huge quantities of structured census data are relatively easy to process using traditional tools.

Instead, the use of the term "Big Data" seems to point toward *qualities* of the data, associated to a *deep failure of traditional approaches*. In other words,

the impact of Big Data is not seen as merely methodological—rather, in the words of Boyd and Crawford (2012), they are associated to "a profound change at the levels of epistemology" p.665.

When it comes to the methods that structure the data, the transition is perhaps best headlined as a shift from mathematically organized data to algorithmically organized data (see also Törnberg, forthcoming). While survey data is constructed for processing through variable-based analysis, requiring pre-compartmentalized data designed to be palatable for a scientific perspective that sees the social world through a lens of averages and variances, data extracted from digital technologies tends to be structured by and for algorithmic processing, implying indexed data structures and traversable networks (Mackenzie, 2012; Marres and Weltevrede, 2013). New data therefore tends to be poorly suited for statistical analysis; it often comes in small chunks, spreading and diffusing in complex and constantly transforming networks, without clearly defined bounds. The social ontology that digital technologies operationalize is not focused on the summing up of a population in fixed categories, but rather on the individuals and their dynamic connections and interactions (Castellani, 2014; Uprichard, 2013). This implies no longer producing data by departing from the aim of a whole, implicitly assumed to be the sum of its parts, but rather departing from the parts and their location within a data structure.

This has been taken to suggest that while census data is produced for scientific analysis, Big Data is a "naturally occurring by-product" (Edwards et al., 2013; Kitchin, 2014), constituted by traces of ongoing social processes rather than something produced for scientific consumption. This ostensible rawness is taken to mean that the ontology revealed by Big Data is in fact the true, relational nature of the social world that had previously been concealed by the survey method. As Allen Barton (1968: 1) puts it, "the survey is a sociological meat grinder, tearing the individual from his social context and guaranteeing that nobody in the study interacts with anyone else in it".

This idea of rawness is seen within the emerging Computational Social Sciences as providing the foundation for a new approach to studying the social world, which is argued to have the potential to solve many of sociology's deep-rooted problems. Alex Pentland refers to this as "sociology of the 21th century" in Manovich, 2011:464, and says that digital data "give us the chance to view society in all its complexity, through the millions of networks of person-to-person exchanges". The new data allows us to navigate "data sets without making the distinction between the level of the individual component and that of the aggregated structure" (Latour et al., 2012: 590), which, Lazer et al. (2009)

argues, has the potential to transform our understanding of our lives, organizations, and societies. Other scholars have argued that "a new kind of social science" is needed (e.g. Christakis, 2012) (a call referring Wolfram's (2002) declaration of "a new kind of science", i.e. Complexity Science), to respond to the fundamental changes that Big Data brings in its wake. This new science is seen as the answer to the crisis of the old approach of empirical sociology, through a supplanting of surveys and interviews with data mining and GIS analysis (Savage and Burrows, 2007). This is seen as bringing a redrawing of the disciplinary boundaries by, as Watts (2007) argues, resolving the issues that has made the social sciences "less successful" than the physical and biological sciences in providing explanatory and coherent theoretical accounts of, for example, the complexities of collective social behavior (see e.g. Bajec and Heppner, 2009; Dorigo and Stützle, 2010; Helbing et al., 2005; Johnson, 2002). Thus, the new data is seen as enabling a convergence between the social and natural sciences under a new approach and ontology (Christakis, 2012). The difference in social ontologies that are operationalized by the old and new data in many ways corresponds to the contrast between "complicated" and "complex" systems (Morin, 2007). This similarity is not incidental: the Santa Fe school of Complexity Science developed largely around the study of, and with, computers and algorithms, in which the dynamics of computer models of mass-interactive systems were studied under labels such as ALife, Agent-Based Modeling and Cellular Automata (Galison, 1997). It was largely this study that became the foundation for the theory of complexity, describing both an ontological category and an approach (e.g. Mitchell, 2009). It is for this reason to little surprise that Big Data seems similar to this ontological category and responds well to a similar methodology, as they carry the same basic social ontology within their structure, having been shaped by the same type of methods and technologies.

What, then, is implied by this ontology of complexity? According to Complexity Science, *complicated* systems have large attribute-rich components, with simple and limited interaction, while *complex* systems typically have many, simple components interacting in sophisticated ways (Andersson and Törnberg, 2017). If the structure of the components of an automobile is an example of the former, the fluid organization of a flock of birds is an example of the latter. Because of the structure of digital trace data, the social ontology of complexity has increasingly become the implicit or even explicit foundation for many of the empirical computational approaches to social science, and the relation- and interaction-focused field of Complexity Science has become a powerful impetus in the development of the

new social scientific disciplines around Big Data (see e.g. Conte et al., 2013; Jungherr, 2015).

The complexity approach has proven highly capable of analyzing many types of systems that have otherwise been impenetrable to formal approaches (e.g. Mitchell, 2009). Complexity Science is centered on the core use of formal models of mass-interaction, focusing not as much on social facts and aggregate explanations, but rather on the emergence of aggregate pattern. This means putting the finger exactly on the limits of aggregate measures, since *emergence* is the very opposite of aggregation (e.g. Wimsatt, 2007: 274–276); *the whole is different from the sum of its parts* (see Anderson, 1972).

For a sociologist, this view of the social world is more reminiscent of Tarde's notion of *imitation* than Durkheim's concept of *social facts* (Candea, 2010; Törnberg, 2017), but while Tarde departed from theory and was criticized for lack of method, Complexity Science developed largely because of, and on the foundation of, new methods (Helin et al., 2014).

The epistemological perspective of complexity relates to the generally accepted view in the various sciences dealing with complex collective behavior, that there exist some fundamental differences between the individual and the aggregate levels (Calhoun, 2002; Knorr-Cetina and Cicourel, 1984). Traditionally, in the social sciences, the existence of levels have often been assumed, and questions have focused mainly on issues like whether the micro or macro level is the suitable level of analysis, or if the two would be possible to "reconcile" using some higher-level theory. In practice, this has primarily been handled through disciplinary and methodological separations, leaving the question of the emergence of structure from individuals to the road-side. Complexity Science, in contrast, focuses explicitly and almost exclusively on this question (Érdi, 2007; Mitchell, 2009).

The Computational Social Science introduced by e.g. Lazer et al. (2009) constituted to certain degree a reboot or re-appropriation of the term: the Computational Social Science of the previous decade was part of Complexity Science and was never linked to large-scale data, but rather approached society mainly through simulation in general, and agent-based modeling in particular. While a certain rift between the data-focused and the simulation-focused Computational Social Science can be identified, they were, and are, strongly connected through a common perspective on society as a relationally and dynamically complex system.

With the rise of Big Data, Complexity Science seems to be increasingly experiencing an "obliteration by incorporation" (Merton, 1968: 27), as the perspective transitions from an explicit focus on emergence and complexity to constituting an implicit foundation in many of the tools and approaches used for social scientific research. For instance, complex network analysis has become widely used within parts of mainstream social science, focusing on how relations/interactions on the micro-level lead to the formation of higher level social patterns (e.g. Strogatz, 2001). The influence of complexity thinking, and the linking of macro dynamics to individual behavior, is also exemplified by many theories and concepts that are increasingly used by social scientists even outside of digital methods, including terms such as "threshold effects" or "tipping-points", "power-laws", "preferential attachment", and so on.

The increasing prevalence of the complexity perspective is often argued to be not only the result of differences in the "rawness" of the data and what it can reveal of the social world, but also to reflect actual changes in the nature of social interaction. There may be certain merits to such claims: just as researchers are shaped by the social life of Big Data, so are its users. Differences that are often emphasized is that digital social life seems more quantitative, regular, predictable (as illustrated by the successes of platform and data analysis companies that subsist precisely on predictive analysis), which is argued to motivate a more natural scientific approach to the data. Two such lines of argument in particular can be identified.

First, social media is seen as having brought with it increased quantification of social interaction, since, for instance, the number of likes that a post has received on Facebook requires no additional operations to be quantified. While traditional data on social interaction requires transcription by researchers, a process that lifts difficult questions about the interpretation of intonations, pauses and subtle facial expressions, the users of platforms like Twitter and Facebook seem to have already done the work of encoding their messages into a quantified and standardized format. In other words, our social life has thus become more natively quantified, more ordered and structured, as we increasingly use numbers and codified data to navigate our social world, flattening the exchange of meaning into numbers, written words, and a choice from a predefined set of smileys. Such native quantitativity has historically implied a tendency for an increased focus on quantitative approaches also for the study of these systems, as is exemplified by the strong mathematical orientation of mainstream economics (Sayer, 1992).

Second, social media is seen as having brought with it an increased prevalence of synchronistic behavior, taking the form of cascades of similar "viral" actions. This has in part been argued to be the result of the ways that social media are designed to bring forth certain type of behavior, by enticing and bringing to the fore

our reactions and instincts, thus undermining the agency we have over our own minds (Alter, 2017). Our technological sophistication is thus said to have, ironically, brought us away from reflexive agency and closer to reactive, animalistic and instinctual behavior well-described by analogies such as "avalanches and granular flows, flocks of birds and fish" (Ball, 2012: ix). The predictability of such online behavior is argued to have enabled the successes of platform and data analysis companies that subsist precisely on predictive analysis.

In summary, the combination between a social life that is more reactive, instinctive and natively quantified, and an understanding of Big Data as something fundamentally *new*, *raw* and *natural*, is stirring to life again the old corpse of naïve naturalism, whose historical refusal to lie down was noted already by Bhaskar (1978). Because of the methods and algorithms that have molded this new digital data, this naturalist ontology takes the shape of *complexity*. This idea of a blurring of the boundary between the natural and social world, and suggested taming of society's "vexatious nature" (Dahrendorf, 1968: 23), may not always be explicit and openly articulated, but is nonetheless apparent in the way a majority of the scholars within the computational approach the social world.

In the following sections, we will look at *the limits of this new naturalism*, by scrutinizing its implicit assumptions about the social world. We will approach these limitations through a contrast with the discussions at the early era of digitalization.

Early digital research and beyond: Liquidity and postmodernity

In the early discussions on the implications of digital technology, pre-dating the age of ubiquitous social media and digital platforms, digital technology was primarily viewed through the lens of dematerialization: the transition of technology *from atoms to bits* (Mitchell, 1996; Negroponte, 1996). The focus in this literature was on the social implications of the possibilities for rapid change brought on by digital technology: through the Internet, technological changes can be distributed to billions of users within seconds, and the reactions of these users can be instantly evaluated. According to these early scholars, the digital is not limited by the constraints of the material world: it has left behind the sculpturing of hard matter for the fluidity of electrons and software. Through this, technology was argued to having reduced its function as a stabilizer of social structures, which implies that the social context and the basis for interpretation become more fluid.

In these discussions, the notions of digitalization and dematerialization were connected to the larger contemporary discussions around terms like "postmodernity", "liquid modernity", "late capitalism", and "acceleration". Analyses of the implications of digital technology can be found in a range of strands, from Baudrillard's (1994) *simulacra*, Jameson's (1991) cultural analysis of late capitalism, Beck's (1992) portrayal of destructuration in the *Risk Society*; in Giddens' (2002) imagery of a disordered *runaway world*; in Bauman's (2000) *liquid modernity* in which "flows" replace the determinate social structure and cultural systems; in Archer's (2014) *morphogenic society* where morphogenesis increasingly dominated over morphostasis.

The common denominator of these views is in many ways the precise opposite of the epistemological conclusions taken in the current debate on digital media: these scholars saw the digital technology as being part of a late modernity "uncontrollable and quintessentially kaleidoscopic in form" (Archer, 2014: 1). As Archer emphasizes, this means that just because a social phenomenon (institution, role, group, belief or practice) continues to bear the same name, "it cannot automatically be regarded as being 'the same'" (p.6), and continuously stable. Digitalization and dematerialization were thus seen as part of the processes of postmodernism in that it constitutes the dissolution of an impediment to the pace of change. This is part of the larger process of modernity, in which, "instead of inhabiting a stable world of objects made to last, human beings found themselves sucked into an accelerating process of production and consumption". (Arendt 1958: xiv)

According to these scholars, digitalization can thus be understood as yet another step or phase of this transition, in which capitalism has, as Jameson (1991) argued, reached its purest form. Through digitalization, this process has finally melted the very materiality of technology, permitting *all that is solid to melt into air*; or, in this case, source code. "It is as though we had forced open the distinguishing boundaries which protected the world, the human artifice, from nature […] delivering and abandoning to them the always threatened stability of a human world" (Arendt, 1958: 126). In this perspective, the stability of the social world is connected to the very materiality of technology: since material change tends to be slow, technologies have provided a relatively solid foundation for social patterns to lean on (Elder-Vass, 2017). For instance, a building can remain standing for hundreds of years and contribute to propagate the social context in which it was constructed. Thus, in the understanding of this literature, it is precisely this stabilization that is undermined by the dematerialization brought by digital technology.

As Hayles (1999) points out, this new instability is brought into our very language, and the ways we interpret the world. To analyze this, Hayles build upon Lacan's concept of "floating signifiers", adding that they through digital technology also begin to *flicker*: our words become unstable, their meaning amorphous and constantly transforming. In Hayles (p.52) words, information technologies "fundamentally alter the relation of signified to signifier[,] thus carrying the instabilities implicit in Lacanian floating signifiers one step further."

Taken together, the view proposed by these early critical scholars can be understood as digitalization bringing increasing *openness* (in the sense of "open systems" in e.g. Bhaskar, 1978) to society as a system by enabling rapid technological change, in turn bringing what Lane and Maxfield (2005) call "ontological uncertainty": an increased propensity for qualitative change. This propensity is illustrated by concepts like "web time" (Karpf, 2012), that describes the increased pace of sociotechnical change brought by information technology. Or in the terminology of Simon (1962), digitalization implies that the "short run", in which a system can be understood formally, becomes increasingly short (Andersson et al., 2014). As Sayer (1992: 122) points out, this in turn limits the usefulness of quantification, since the objects under measure are not qualitatively invariant.

In our view, despite being largely neglected in the contemporary literature, this early characterization of digitalization remains in many ways accurate as a description of the effects of digitalization on social life; implying a fluidity and instability of meanings and structures constantly boiling under the surface of the ostensible constancy of fixed numbers and symbols. However, digital technology has since developed in some new and at the time unforeseen directions, which have meant that the fluidity of meaning and structures have become channeled in unexpected ways.

Fluid technology in the era of platforms

First, at the time of these theories, the Internet was a highly fragmented environment of rapid and often informal experimentation. Today, the Internet has infrastructurally instead become a place of extreme centralization: information systems have turned out to be ripe with natural monopolies, creating conditions for large platform companies. The structures that typically follow from this have become increasingly similar to a form of private governments, with power to control flows of information, and, as the "sharing economy" (e.g. Uber and Airbnb) illustrates, at times even to tax their user base.

Second, it is not only the roll-out of new technology that has changed with digitalization, but also the broader feedback processes of innovation, in particular the evaluation of how new innovations affect the social web in which they become part. Sophisticated data analysis, A/B testing, and instantaneous evaluation of the social practices evolving on digital platforms enable platform owners to shape their users' behavior with unprecedented precision and control. The feedback loop between evaluation and innovation (described by e.g. Lane, 2016) has become increasingly rapid, as technology owners have precise and detailed data on how their products are taken up in a larger sociotechnical context.

These two factors have meant that the fluidity and capacity for rapid change of dematerialized technology, theorized by the scholars of the early days of the Internet, have not only played into a postmodern culture of late capitalism, but has also been channeled into new forms of power for the owners of technology. Technological power can now be exercised in more sophisticated, nimble and illusive ways than ever before, as the dematerialization of technology means that the ownership even of consumer products has become possible to centralize. The artifacts that we consume and surround ourselves with are increasingly rented rather than owned, as apps, programs, and technological platforms are increasingly located in the cloud, and thus prone to constantly change without warning. The "zero-marginal-cost" of software has resulted not in the end of capitalism, as some social scientists rather optimistically theorized (Mason, 2015; Söderberg, 2015), but rather in a transition of business models from *selling* to *renting*. In other words, rather than undermining the private ownership regime of capital, this has had the effect of undermining the already tenuous ownership of consumers (von Busch, 2008). Instead of workers gaining the ownership of the means of production, they have increasingly lost ownership even of their goods of consumption.

While digitalization brings increasing centralization and sophistication in the expression of technological power, technology's function as a shaper of social behavior is in itself nothing new. Technology has always been in and for the power of its owners and producers, as a force capable of shaping and directing social life in their interests. There is hardly an activity, belief or form of interaction that is not mediated by artifacts and thus affected by this hidden ideological face of technology (Feenberg, 1991), whether wedding rings and clothes, candles and incense, or money and art—these artifacts store and propagate societal structures (Elder-Vass, 2017). Social life has always played out within technological platforms that shape and frame our interaction and provide context to it, granting permanence to our symbols and our language (Collins, 2014). The impact of the technological context

is not merely incidental: churches, for instance, are expressions of power and authority, consciously designed to inspire awe toward the power of Gods, religious institution and holy men. They instill authority into the solemn priest behind the pulpit, and remind us of the larger story within which we are but minor players, thus shaping and giving meaning to our behavior and interaction. Online digital platforms of today are not unlike such physical meeting places: they too provide the context within which rituals and social life take place. They condition our interactions, shape who has authority and who is heard.

But the combination between centralization of power and the dematerialization of technology implies an important transition in the expression of technological power. While yesteryears churches were carved from stones, rocks and clay, the digital churches of today constantly shift underneath our feet. While physical churches were blunt tools for shaping our lives, needing to be backed up by damnations and inquisitions, the digital churches read and react to our every gesture and expression. They are capable of customizing their expressions to individuals, or trying a hundred variations of the colors of the pulpit to see how its faithful are affected. What use culture emerges among users is importantly controlled by what the system "affords" (Norman, 1999), and what can be done "frictionlessly" (Shaw, 2015): subtle design choices herd the users in certain directions, in ways related to the concept of "nudging" (Thaler et al., 2013).

In other words, the increasing fragmentation and fluidity following from dematerialization has somewhat paradoxically implied increased centralization of control, as it permits the owners of technology to express power by shaping meaning and structures through gentle nudging of underlying technical rules. This control does not congeal the constant boiling fluidity of meaning, but rather dynamically directs its flow. Control moves to lower ontological strata, shaping outcomes through the underlying rules of interaction rather than through explicit control. In this dematerialized modernity, the fluidity of meanings and structures afford a form of control that seemingly paradoxically emerges from the bottom up.

This transition in the expression of power is reminiscent of the transition described by Norbert Elias (2006) in *The Court Society*. Just as Luis XIV embedded his control into the social rules of polite interaction rather than, as previous regimes, through violence and explicit control, power in the era of digital platforms is expressed not top-down, but through invisible nudging and shaping of local behavior, molding of social rules and practices, and thus, control is embedded in the very rules of our interaction. The interests of platform owners thus appear to us as

seemingly natural and spontaneous outcome of human behavior. This form of distributed control fits into the individualization of power that is part and parcel of postmodernity; as Bauman (2000) notes, control has become part of individuality itself; no longer is the focus on producing homogeneity by whipping deviators into conformity, but rather on the emergence of a collective outcome in line with certain interests—directing the herd rather than the beast, through a shaping of context rather than through explicit command-and-control. The transition from technology being a rather blunt tool for social control to a virtual social scalpel thus implies that digitalization has brought an era of *platform power*, in which technology provides a new level of herd control.

The nature of digital data

What, then, are the implications of the condition of postmodernity and the technological power of platform owners for digital data research? What limits do these observations imply for the computational study of digital social life, and how do they clash with the tendency of Complexity Science to naturalize social life—seeing social patterns not as the result of contingency and conflict, but as expressions of universal social laws?

As we saw in the first section, the promise of the Big Data revolution has described a world of previously unimaginable data; a flood of coffee-table discussions revealing traces of the lives, dreams, and feelings of hundreds of millions of people. This has painted a picture—which hangs centrally in the halls of e.g. Computational Social Science—of the "true" relational nature of social life being unveiled, showing a social life which is not only measurable but even predictable.

While this painting shows a dreamy world for the social sciences, another reality appears when we lift our gaze from the data feeders, and cast it upon the less than appetizing context in which the data is fed to us. Rather than a spontaneous and natural production of social traces, we see how the data is produced, selected and provided to us by platform owners pursuing their own interests. Many aspects are left out of this data. For instance, the platforms and their rules that shape the online behavior are not readily visible: their interests and incentives instead lie latent as hidden forces that guide individual behavior and the emergent social practices of the platforms. Thus, at the same time as the contextual aspects and the power of platform owners are becoming increasingly central to understanding social life, our focus as researchers is increasingly on the patterns of interaction, which, as they lose their natural setting, become naturalized and decontextualized, just in the way that complexity perspectives have historically had a tendency of implying

naturalization (Byrne and Callaghan, 2014; Uitermark, 2015). When Big Data is seen as merely an *encoded*, measurable version of social reality—possibly with some technological bias to be corrected for—the complex social and technological forces that produced them are flattened: the data is made to seem natural and inevitable rather than contingent and contested; they are made subject of reification rather than critique.

This idea of Big Data as an "encoding" of social life disregards the complex interplay between the technological and social aspects of human life that have produced the data. Rather than merely a one-way encoding, the production of data takes place by digital platforms directing and limiting action by providing a "grammar of action" that make certain activities doable, and thus rendering social activities available for measurement, analysis, commodification, and manipulation (Van Dijck, 2013). But at the same time, users are not helpless puppets in this process: they are often aware of the ways that measures and technologies play into their social lives, and reflexively take account of this in their use. They are not "encoding" their behavior, but rather employing and enacting the methods, performing through the measures in front of an "imagined audience" (Litt, 2012): "social actors produce methodical accounts of social life as part of social life" (Lynch, 1991). The platform owners are in turn aware of these dynamics: the very creation of digital platforms tends to involve the implementation of sociological and social psychological ideas; their use ranging from the benign push (e.g. suggesting friends through triadic closure) to what basically amounts to a weaponized social psychology (e.g. "engagement maximization" through the application of research on addiction). In short, measures not only describe, but are enacted and made part of social life, in the type of continual process of reflexivity between diverse actors and roles that is quintessential of the vexatious nature of social life.

Online behavior and content are in other words a consequence both of how digital technologies work and what people do with them, in ways that are exceedingly difficult to separate. Rather than thinking of online social life through a separation between "human behavior" to be studied, and "technological bias" to be in various ways "corrected for", content is perhaps better understood as the output of an entanglement between the two—a sociotechnical system (Marres, 2017). This casts technology as a defining feature of human society, rather than as something to be corrected for. The way that "virality" is employed to make claims about the new "instinctual" and "reactive" nature of digital social life is the case in point here; Halavais (2014) shows how the re-tweet emerged as a sociotechnical script on Twitter: beginning as an informal practice, to becoming encoded in a button, which ended up producing the macro-pattern of "virality" and "diffusion", appearing as repeated behavior cascading through a network.

The result of thinking of Big Data as providing a form of privileged access to the social world is that researchers flock to study relatively predictable and correlated social behavior on ostensibly disintermediated online platforms, while disregarding the sociotechnical conditions that lead to the formation of that behavior. The digital platforms are developed using significantly more sophisticated methods and larger data quantities than what is available to researchers, with even the most seemingly insignificant design decision being the result of meticulous A/B-testing and data analysis. From the basis of the Complexity Science metaphor of social behavior as the playing of a game, the data thus makes more visible the "playing of the game", while obscuring the "rules of the game" and the interests that shaped them. The data thus becomes a perfect fit for a naturalizing science that tends to see the rules as universal and their outcomes as inevitable.

The formal tools and mathematical models that we apply to study this world hinge on the stability of meaning and understanding that are exchanged. But such assumptions have not become less problematic through digitalization, but rather more so, as symbols and meaning are becoming more local in time and space. Interpretation has not become less central in the research process; its locus has merely moved, as interaction is simultaneously more quantified and its meaning more fragmented and flickering.

This does not mean that the observation of increasing complexity, native quantitativity, and the potential for predictability are false. Big Data is seemingly paradoxically associated to both these developments: it is simultaneously more liquid *and* more natively quantified; it is simultaneously more open *and* more measurable; it is simultaneously more bottom-up and more amenable to control. It is, in short, becoming easier to count, while at the same time harder to interpret what we are counting.

The answer is not, as has been the case among some scholars, to simply reject computational methods or suggest that the entire notion of "new data" is merely a red herring since many of its aspects have a long history (Marres, 2017; Uprichard et al., 2008)—the epistemological and methodological demands of complexity in general, and Big Data in particular, are real and will have to be reckoned with. But neither is the answer a methodological one: we *will* not find any method to match and capture society in a single analogy (Andersson et al., 2014; Archer, 2014). The solution needs to concern the underlaborer on which the approach is founded. Instead of continuing to

approach Big Data by extending "the tool found successful in one domain to decipher the other" (Khalil, 1995: 414–415), we suggest following Perona's (2007) advice with regards to social complexity: to take "a turn to ontology". Ontological perspectives are not only matters of philosophical curiosity, but have profound implications for how we can and should research, manage, and think about social phenomena. What is needed is a meta-theory capable of respecting the openness and non-decomposability of social systems in general, and digital social systems in particular, while at the same time admitting the methodological and epistemological conditions of Big Data—large data sets characterized by relational complexity, emergence, and self-organization. Complex Realism provided such a response to the insights of Complexity Science in the 90s, bringing the patchy and partial social ontology of complexity into dialogue with Critical Realism (e.g. Byrne and Callaghan, 2014). A potential way forward, to be pursued in future publications, could thus be to follow an analogous response to the insights of digital data.

Conclusions

The first section of this paper showed how the structure of digital data is making trouble for the traditional social scientific variable-based approach, creating a push toward new social ontologies matching the structure of the data. This has sparked a renewed, *complex* naturalism, within which social systems are increasingly approached through the formal methods of the natural sciences—seeing social structures as patterns naturally emerging from mass-interaction, which is taken to permit the leaving out of institutional, technological, and contextual aspects of social life. In the second section, we revisited the discussions of the early days of digitalization: these instead saw digitalization as part of the larger processes of postmodernity, implying increased systemic openness as the transition from atoms to bits brought the undermining of stabilizing forces of social systems. We extended this perspective by discussing how the liquidity of technology has turned into a means of control as online social life has become centralized into large platforms that work to shape human behavior according to their interests. Together, these developments reveal a clash between the underlying assumptions of the computational approach to social data and the context in which the data is produced. While the trouble-making of digital data may usefully help point to limits of the traditional variable-based approach as well as the constructed nature of scientific data, the new data brings new limits, and are similarly constructed around certain methods and techniques. The early debate on

digitalization thus serves as a reminder of aspects of the social world that the new computational view continues to leave out: as Andersson et al. (2014) argue, society is neither a *complex* nor a *complicated* system, but rather, it displays *both* these properties which makes it qualitatively different from both types of systems: thus, the reduction of social reality to these "analogical imaginations are simply misleading" (Archer, 2013: 146).

The technological power of platform owners is to a large part enabled by the same new tools for data analysis as used by social scientists—indeed, the private sector is often the driver for the development of these tools. These efforts have been immensely economically successful, as illustrated by companies like Google and Facebook. But we must not forget that the aims of these corporations are quite different from the aims of researchers: they seek *prediction* and *control*, while researchers (at least should) seek *explanation* and *understanding*. In trying to make a user click an ad, corporations are less interested in the *why* than the *how*. These aims are implicitly built into the affordances of the tools, and just like the online platforms shape the actions of their users, these data analytical tools tend to shape the behavior of their users, that is, our behavior as social scientists. They thus nudge researchers toward pattern-finding and prediction, rather than in-depth understanding.

As noted by early scholars of digital technology, the flexibility and rapid change enabled by digitalization are part of the larger processes of postmodernity. But the digital world is not well-described by the classical understanding of postmodernity alone. It is also part of an increased centralization of technological power, and a change in the role of technology in social life. The postmodern aspects of digitalization do bring more "openness" (in the sense of e.g. Bhaskar, 1978) to society, with social structures becoming more fragmented, *liquid*, and prone to qualitative change, which can arguably be seen in some cultures developing on the Internet (Nagle, 2017). This has furthermore rightly been understood to pose limits on quantification, since it implies that the measured objects are not qualitatively invariant (Sayer, 1992, 2000).

This is made more confounding by the fact that this development is occurring in an increasingly natively quantitative context, in which people are communicating through numbers and coded messages. While this changes the locus of scientific interpretation, since researchers no longer need to transcribe conversations, it does *not* reduce the centrality of interpretation. While transcription brings the researcher a direct experience of the inherent loss of nuances of meaning, making the local and contextual nature of meaning hard to ignore, the digital platforms conceal the ways that their

quantitative data is produced by fluid and quickly changing sociotechnical systems, whose signifiers flicker and vary over time and context. These changes are determined in part by constantly evolving implicit social practices that are self-consciously hyper-ritualized, fragmented, and local to social context, and co-evolving with technological change in the underlying platforms.

Similarly problematic is the notion that human online behavior being more reactive should lead us to disregard context and view social life through analogues like "flocks of birds and fish". While technological platforms do act on humans through their powerful social life, wielded and directed by platform owners, the fact that platforms are capable of herding users through technological nudges and affordances should not imply a reduced focus on context, but rather engage us to turn our gaze toward precisely the power of the platforms. But instead of moving towards increasing focus of these contextual aspects and the role of platform owners, many scholars interested in digital social life have been lured by the siren-call of new methods and abundant digital data to lose their gaze from precisely these factors. Thus, social patterns are naturalized: we take behavior on social platforms to be telling about the nature of social life, while it may in fact say more about the interests of the platform owners.

For us as researchers, this implies a need for not only studying processes of emergence, but for doing so while keeping in mind that there is nothing natural about human behavior and that there is no such thing as "raw data". Social media should perhaps be thought off less like savannahs of free-running herds of humans, than like zoos in which caged users are made to dance to the tune of capital; "no data, big or small, can be interpreted without an understanding of the process that generated them" (Shaw, 2015: 1), and these processes are entangled in the interests of capital.

This calls for a *critical* computational social science that does not sacrifice context, clarity, and critique for the automatic identification of large-scale patterns, predicated in the notion that breadth could replace depth and context as basis for interpretation. If we are not to be drawn by the siren-song of abundant data, sung by the owners of technological platforms precisely to lure us into drowning in the data deluge, we must tie ourselves to the mast of a critical and explicit metatheory: for only from a stable ontological position will we be able to hear not only what the data has to sing to us about the social world, but also to listen for those things about which they remain so curiously silent.

ORCID iD

Petter Törnberg http://orcid.org/0000-0001-8722-8646

References

Alter A (2017) *Irresistible: The Rise of Addictive Technology and the Business of keeping us Hooked*. London: Penguin.

Anderson C (2008) The Petabyte age: Because more isn't just more–more is different. *Wired Magazine* (16.07): 106–120.

Anderson PW (1972) More is different. *Science* 177(4047): 393–396.

Andersson C, Törnberg A and Törnberg P (2014) Societal systems – Complex or worse? *Futures* 63: 145–157.

Andersson C and Törnberg P (2017) Wickedness and the anatomy of complexity. *Futures* 95(1): 118–138.

Archer M (2014) *Late Modernity: Trajectories towards Morphogenic Society*. Cham: Springer Science & Business Media.

Archer MS (2013) *Social Morphogenesis*. Dordrecht: Springer Science & Business Media.

Arendt H (1958) *The Human Condition*. Chicago, IL: University of Chicago Press.

Bajec IL and Heppner FH (2009) Organized flight in birds. *Animal Behaviour* 78(4): 777–789.

Ball P (2012) *Why Society is a Complex Matter: Meeting Twenty-first Century Challenges with a New Kind of Science*. Cham: Springer Science & Business Media.

Barton AH (1968) Bringing society back in: Survey research and macro-methodology. *The American Behavioral Scientist* 12(2): 1–9.

Baudrillard J (1994) *Simulacra and Simulation*. Ann Arbour: University of Michigan Press.

Bauman Z (2000) *Liquid Modernity*. Cambridge: Cambridge Press.

Beck U (1992) *Risk Society: Towards a New Modernity*. London: Sage Publication Ltd.

Bell G, Hey T and Szalay A (2009) Beyond the data deluge. *Science* 323(5919): 1297–1298.

Bhaskar R (1978) *A Realist Theory of Science*. Hassocks: Harvester Press.

Boyd D and Crawford K (2012) Critical questions for big data: Provocations for a cultural, technological, and

scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679.

Byrne D and Callaghan G (2014) *Complexity Theory and the Social Sciences: The State of the Art*. London / New York, NY: Routledge.

Calhoun C (2002) *Classical Sociological Theory*. Oxford: Blackwell.

Candea M (2010) *The Social after Gabriel Tarde: Debates and Assessments*. New York, NY: Routledge.

Castellani B (2014) Focus: Complexity and the failure of quantitative social science. *Discover Society*. Available at: https://discoversociety.org/2014/11/04/focus-complexity-and-the-failure-of-quantitative-social-science/ (accessed September 2018).

Christakis NA (2012) A new kind of social science for the 21st century. *Edge*. Available at: http://edge.org/conversation/a-21st-century-change-to-social-science (accessed September 2018).

Collins R (2014) *Interaction Ritual Chains*. Princeton: Princeton University Press.

Conte R, Hegselmann R and Terna P (2013) *Simulating Social Phenomena*. Heidelberg: Springer Science & Business Media.

Dahrendorf R (1968) *Essays in the Theory of Society*. Stanford, CA: Stanford University Press.

Dorigo M and Stützle T (2010) Ant colony optimization: Overview and recent advances. In: Gendreau M and Potvin JY (eds) *Handbook of Metaheuristics*. *International Series in Operations Research & Management Science*. Vol. 146. Boston, MA: Springer.

Edwards A, Housley W, Williams M, et al. (2013) Digital social research, social media and the sociological imagination: Surrogacy, augmentation and re-orientation. *International Journal of Social Research Methodology* 16(3): 245–260.

Elder-Vass D (2017) Material parts in social structures. *Journal of Social Ontology* 3(1): 89–105.

Elias N (2006) *The Court Society*. New York, NY: Pantheon Books.

Érdi P (2007) *Complexity Explained*. Berlin: Springer Science & Business Media.

Feenberg A (1991) *Critical Theory of Technology*. New York, NY: Oxford University Press.

Galison P (1997) *Image and Logic: A Material Culture of Microphysics*. Chicago, IL: University of Chicago Press.

Giddens A (2002) *Runaway World: How Globalisation is Reshaping Our Lives*. London: Profile books.

Halavais A (2014) Structure of Twitter: Social and technical. In: Weller K, Bruns A, Burgess J, et al. (eds) *Twitter and Society*. New York, NY: Peter Lang, pp. 29–42.

Hayles NK (1999) *How We Became Posthuman*. Chicago, IL: University of Chicago Press.

Helbing D, Buzna L, Johansson A, et al. (2005) Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions. *Transportation Science* 39(1): 1–24.

Helin J, Hernes T, Hjorth D, et al. (2014) *The Oxford Handbook of Process Philosophy and Organization Studies*. Oxford: Oxford University Press.

Hey T, Tansley S and Tolle Kristin M (2009) *The Fourth Paradigm: Data-intensive Scientific Discovery*. Vol. 1. Redmond, WA: Microsoft research.

Jameson F (1991) *Postmodernism, or, the Cultural Logic of Late Capitalism*. London: Verso.

Johnson S (2002) *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*. New York, NY: Simon and Schuster.

Jungherr A (2015) *Analyzing Political Communication with Digital Trace Data*. Cham: Springer International Publishing.

Karpf D (2012) Social science research methods in Internet time. *Information, Communication & Society* 15(5): 639–661.

Khalil EL (1995) Organizations, naturalism, and complexity. *Review of Social Economy* 53(3): 393–419.

Kitchin R (2014) Big Data, new epistemologies and paradigm shifts. *Big Data & Society* 1(1): 1–12.

Knorr-Cetina KD and Cicourel AV (1984) *Advances in Social Theory and Methodology: Toward an Integration of Micro- and Macro-sociologies*. Boston, MA: Routledge & Kegan Paul.

Lane DA (2016) Innovation cascades: Artefacts, organization and attributions. *Philosophical Transactions of the Royal Society B* 371(1690): 1–9.

Lane DA and Maxfield RR (2005) Ontological uncertainty and innovation. *Journal of Evolutionary Economics* 15(1): 3–50.

Latour B, Jensen P, Venturini T, et al. (2012) 'The whole is always smaller than its parts' – A digital test of Gabriel Tardes' monads. *The British Journal of Sociology* 63(4): 590–615.

Lazer D, Pentland AS, Adamic L, et al. (2009) Computational social science. *Science* 323(5915): 721–723.

Litt E (2012) Knock, knock. Who's there? The imagined audience. *Journal of Broadcasting & Electronic Media* 56(3): 330–345.

Lynch M (1991) Method: measurement ordinary and scientific measurement as ethnomethodological phenomena. In: Button G (ed.) *Ethnomethodology and the Human Sciences*. New York, NY: Cambridge University Press, pp. 77–108.

Mackenzie A (2012) More parts than elements: How databases multiply. *Environment and Planning D: Society and Space* 30(2): 335–350.

Manovich L (2011) Trending: The promises and the challenges of big social data. *Debates in the Digital Humanities* 460–475. Available at: http://manovich.net/index.php/projects/trending-the-promises-and-the-challenges-of-big-social-data.

Manovich L (2016) Online post. Facebook. Available at: https://goo.gl/GdcJhm (accessed September 2018).

Manyika J, Chui M, Brown B, et al. (2011) *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute. Available at: https://bigdatawg.nist.gov/pdf/MGI_big_data_full_report.pdf (accessed September 2018).

Marres N (2017) *Digital Sociology: The Reinvention of Social Research*. Hoboken: John Wiley & Sons.

Marres N and Weltevrede E (2013) Scraping the social? Issues in live social research. *Journal of Cultural Economy* 6(3): 313–335.

Mason P (2015) *Postcapitalism: A Guide to Our Future*. London: Penguin Random House.

Merton RK (1968) *Social Theory and Social Structure*. New York, NY: Free press.

Mitchell M (2009) *Complexity: A Guided Tour*. Oxford: Oxford University Press.

Mitchell WJ (1996) *City of Bits: Space, Place, and the Infobahn*. Cambridge: MIT press.

Morin E (2007) Restricted complexity, general complexity. In: Gershenson C, Aerts D and Edmonds B (eds) *Worldviews, Science and Us: Philosophy and Complexity*. Singapore: World Scientific.

Nagle A (2017) *Kill All Normies: Online Culture Wars From 4Chan and Tumblr To Trump And The Alt-Right*. Winchester: Zero Books.

Negroponte N (1996) *Being Digital*. New York, NY: Knof.

Norman DA (1999) Affordance, conventions, and design. *Interactions* 6(3): 38–43.

Perona E (2007) *The Confused State of Complexity Economics: An Ontological Explanation*. Milano: Springer.

Rosling H (2010) *The Joy of Stats*. London: BBC4.

Savage M and Burrows R (2007) The coming crisis of empirical sociology. *Sociology* 41(5): 885–899.

Sayer A (1992) *Method in Social Science*. 2nd ed. New York, NY: Routledge.

Sayer A (2000) *Realism and Social Science*. London: Sage Publication Ltd.

Shaw R (2015) Big Data and reality. *Big Data & Society* 2(2): 1–4.

Simon HA (1962) The architecture of complexity. *proceedings of the American Philosophical Society* 106(6): 467–482.

Strogatz SH (2001) Exploring complex networks. *Nature* 410(6825): 268–276.

Söderberg J (2015) *Hacking Capitalism: The Free and Open Source Software Movement*. New York, NY / London: Routledge.

Thaler RH, Sunstein CR and Balz JP (2013) Choice architecture. In: Shafir Eldar (ed.) *The Behavioral Foundations of Public Policy*. Princeton, NJ: Princeton University Press.

Törnberg A (2017) *The wicked nature of social systems – A complexity approach to sociology*. Gothenburg University, Gothenburg.

Törnberg P (forthcoming) Complex realist economics: Toward an ontology for an interested pluralism. *Review of Social Economy*.

Uitermark J (2015) Longing for Wikitopia: The study and politics of self-organisation. *Urban Studies* 52(13): 2301–2312.

Uprichard E (2013) Describing description (and keeping causality): The case of academic articles on food and eating. *Sociology* 47(2): 368–382.

Uprichard E, Burrows R and Byrne D (2008) SPSS as an 'inscription device': From causality to description? *The Sociological Review* 56(4): 606–622.

Van Dijck J (2013) *The Culture of Connectivity: A Critical History of Social Media*. Oxford: Oxford University Press.

von Busch O (2008) *Fashion-able. Hacktivism and engaged fashion design*. University of Gothenburg.

Watts DJ (2007) A twenty-first century science. *Nature* 445(7127): 489–489.

Watts DJ (2013) Computational social science: Exciting progress and future directions. *The Bridge on Frontiers of Engineering* 43(4): 5–10.

Wimsatt WC (2007) *Re-engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge: Harvard University Press.

Wolfram S (2002) *A New Kind of Science*. Champaign, IL: Wolfram media.