

Materialising the Digital

Materialising the Digital seeks to interrogate the infrastructures, relationships and imaginaries of digital technologies through situated, empirical analyses of the production, circulation and use of digital devices.

SERIES EDITORS:

HANNAH KNOX AND ADAM FISH

Positioned at the intersection of media studies, STS, anthropology and sociology, the series will provide original, critical and theoretically innovative understandings of the implications of digital technologies for contemporary social life. Our intention is that this series will provide a solid ground from which to engage and critique the persistence of utopian, functionalist and dystopic visions of technological futures.

Ethnography for a data- saturated world

Edited by Hannah Knox and
Dawn Nafus

Manchester University Press

1

Introduction: ethnography for a data-saturated world

Hannah Knox and Dawn Nafus

It is increasingly difficult to attend to social and political relations in the contemporary world without recognising that they are in some way constituted by digitally generated data. From censuses that describe national populations to polls that predict and chart election outcomes, from audience surveys and click-counters that are used to price advertising to credit ratings and market analyses that determine financial relations, social worlds are entangled with data that is produced, circulated and analysed using computational devices. To paraphrase Walter Benjamin's famous aphorism about the effects of the then new technologies of film and photography on human engagement with the world, 'every day it seems the urge grows stronger to get hold of a subject at very close range by way of its [data]' (Benjamin 2008 [1939]).

During the 2000s, with the continued increase in computational information processing capacity and the huge spread of smartphones and sensors there has been an increasing public concern about the challenge of data's 'bigness' (Anderson 2008; Bowker 2014). Practices of data collection and collation seem to have exploded in recent years with the proliferation of electronically connected devices that are capable of sensing and producing data about the world and circulating that data to a range of users including governments, corporations and individuals. Using analogies from older industries, the economic and social potential of data has led to its characterisation as the 'new oil', offering potentially new revenue streams, new ways of imagining and

governing populations and new methods of verification and accountability. Those who are more concerned about the political structures and effects of this new resource also talk of data as 'exhaust' – the byproduct of human interaction that needs to be both 'captured' by the analytic converter of data science and properly managed and governed to mitigate the dangers associated with ambiguous attribution, security, corporate monopoly and nefarious techniques of surveillance and control.

Most recently, other social, political and ethical questions have arisen about the implications of automation and machine learning.¹ Newer computational techniques for parsing large datasets focus mainly on what machines can and cannot recognise, asking whether some data has enough of the same features as some other data such that a machine can determine that they are both indeed a picture of a dog, or a stressed tone of voice. These automation practices intensify a sense of opaqueness. Many worry that machine learning systems can grow so complex that it can be difficult even for the very people who designed the system in the first place to say how machines make the determinations that they do. Consideration of the social implications of automation has provided a new realm of debate about data and its ethical implications. No longer are questions about data merely a matter of how objects and subjects become known through different quantities and qualities of data collection, nor are they about who has and should have access to that knowledge. They have been extended to incorporate more fundamental questions about what happens to our sense of what knowledge is when the agents of knowledge production are no longer necessarily even human.

New data relations thus not only raise questions about how to better know and act upon the world, but also shed light on the very foundations of what we consider knowledge to be. This book starts from the conceit that attention to digital data opens up the possibility of interrogating more broadly the presuppositions, techniques, methods and practices out of which claims about the value and purpose of knowledge gain power. To talk of digital data is to talk of one facet of a broader terrain of knowledge production, of which numerical or digital data is only one part. Seeing data practices and concerns as a matter of how to more broadly understand and make the world demands then that we locate digitally collected data as one of many ways of knowing, which include critical

reflection, affective experience and, most importantly for this collection, ethnography.

In spite of the level of enthusiasm and debate about the possibilities and challenges of big data, grounded empirical studies of the knowledge practices entailed in contemporary data analytics are surprisingly few and far between. The journal *Big Data & Society* has done much to generate a social response to big data issues but this is one of very few places where ethnographic accounts of big data as a field of practice exist at all. In part this is no doubt due to the time that it takes for ethnographies to work their way through the publishing system. There are some important studies in the pipeline such as Nick Seaver's (2015) doctoral study on music recommendation analysts and Asta Vonderau's current research project on cloud computing, but at the date of writing these are yet to be published. Meanwhile other data-related phenomena such as practices of modelling and visualisation in scientific settings (Dunitz 2004; Myers 2015) the appearance of bitcoin (Maurer 2012) and the building of databases for the collection and navigation of hybrid and indigenous knowledge forums (Shrinivasan et al. 2009; Verran and Christie 2014) provide an important starting point from which to approach big data practices ethnographically. Such studies are much needed as a way of cutting through the media hype in business press around big data and its promises (see Boellstorff and Maurer (2015) for early work in this area). But ethnographic studies also offer more than just empirical detail that can provide a reality-check on otherwise hyped phenomena. Ethnography done well also holds the promise of generating a new way of theorising and understanding digital data by building novel analytical concepts that are appropriate to the kinds of relations of knowledge production that digital data itself entails.²

This book therefore aims to fill this gap of ethnographic approaches to contemporary digital data by providing a window on to the cultures, practices and infrastructures and epistemologies of digital data production, analysis and use. Understanding the production and use of digital data and its implications for knowledge is an issue that cuts across a huge array of different areas of practice (science, commerce, government, development, engineering etc.) and covering this terrain in its entirety is far beyond the ability of any one volume. In order to provide a path through this complexity we therefore take as our core focus the way in which digital data is troubling and reconstituting

expertise. This focus on expertise allows us to do something that is relatively unusual in an edited collection: both to provide a comparative description of a number of empirical fieldsites where communities of experts are self-consciously forming around the new possibilities put on the table by digital data; and to consider how our understanding of the ways experts make and remake digital data might reframe our own expertise as ethnographers. This is not a methods book, but it is a book about what digital data is doing to empirical methods that sustain claims to expertise, with a particular focus on its implications for ethnography.

We approach digital data then, as a comment on the relationship between knowledge, expertise and the methods through which knowledge is produced. We do this in order to interrogate whether data practices might be part of a broader unsettling of how to know the social. We focus specifically on the interplay between digital data and ethnography as two ways of understanding contemporary possibilities available for knowing, formatting and intervening in the world. This is not just a book about how ethnographic knowledge can fill in the gaps of data science (e.g. boyd and Crawford 2012) nor is it just a demonstration of how ethnography can shed light on what data science actually is and the effects it produces (although both of these are touched upon in this volume). Rather, the book sets out a more ambitious aim of exploring what might be happening to social knowledge production at the interface of data and ethnography, with a view to outlining new directions in social research and simultaneously attending to the epistemological foundations of that research.

Past experiments in digital data and ethnography

The conversations that this book charts between digital data³ and ethnography offer, we suggest, a fresh terrain in which to ask questions about the social production of knowledge. However, the question of how to combine data-oriented and qualitative approaches in ethnographic research is not new.⁴ Anthropologists have, since at least the 1960s, periodically turned to the possibilities that computation might hold for assisting with anthropological analysis. Gregory Bateson and Margaret Mead's forays into cybernetics as a method for analysing social systems offer an early example of how information-theoretical

thinking was incorporated into anthropology and used to reshape a distinctive approach to the discipline (Bateson 1972; Mead 1968). Ecological anthropologist Roy Rappaport's groundbreaking study of the relationship between ritual and ecology offered a similarly systems-theoretical method of socio-natural analysis to chart the relationship between the abundance or scarcity of ecological resources and ritual process, an approach which has more recently been taken up in computer simulation work in Bali by Stephen Lansing and colleagues (Rappaport 1977). Lévi-Strauss meanwhile explored the conceptual potential of computers in the development of structural anthropology and was conversant with the logic of information theory and its influence on structural linguistics (see Seaver 2014b; Geoghegan 2011). Whilst these first theoretical explorations into systems theory and structural analysis took place in the 1950s and 1960s, their influence has gained traction again in recent years and is now felt in much contemporary anthropology, particularly amongst those who study ecological relations and technology (Boyer 2013; Kohn 2013).

As computers developed and became more affordable, a number of anthropologists were quick to explore the broader methodological potential of these new computational devices for assisting with the collection and analysis of field materials. This is outlined in books like Dell Hymes's 1965 volume on the use of computers in anthropology (Hymes 1965). Studies such as Marie Corbin and Paul Stirling's database-supported analysis of kinship and family in Spain in the 1970s established a precedent for the use of computers in anthropological analysis. The Centre for Social Anthropology and Computing (CSAC) was established at the University of Kent in 1986. This remains a key location for discussions and collaborations around the use of computers in anthropological analysis (Ellen 2014).

A parallel field in which anthropologists have played an important part is the study of human-computer interaction. Human-computer interaction (HCI) scholars have a well-established history of entangling digitally produced data with ethnography. During the 1980s for example, anthropologists working at the Palo Alto Research Centre at Xerox Park were noted for bringing ethnomethodological approaches to HCI, which drew on numerical and video data alongside ethnographic fieldnotes to understand the social dynamics of situated computer use (Trigg et al. 1991; Suchman et al. 1999).

Each of these forays into the possibilities that computers might hold for anthropology came at a particular historical moment that brought together specific configurations of people, devices, work practices, questions, theories and intentions. The approach we take in this book is not teleological or historiographic, but rather takes its lead from the contemporary moment, and in particular scholarship that has focused on what has been termed 'the social life of method' (Ruppert et al. 2013; Lury and Wakeford 2012; Marres and Welleverde 2013). This scholarship argues that social science methods are more than just incremental techniques for understanding the world. Methods are also social phenomena in and of themselves, both because they emerge from particular social worlds that organise ontologies and epistemologies in their own particular ways, and because methods actively participate in the social worlds they were designed to comprehend. Surveys were developed as professional instruments for knowing about the concerns of a population, and became the preferred technique of knowledge production for a technocratic middle class who used them to reshape social relations into practices that could be surveyed and audited (Strathern 2000). Ethnography's origins in colonial encounters provide another notorious example. Empire-building set the context in which 'holistic' understandings of subjugated peoples became necessary. State actors mobilised holism in order to shore up notions of 'tribes' as so many distinct units that could be managed and controlled conveniently as single entities, in contradistinction to white settlers. Turn-of-the-century ethnography is not the same as postcolonial or contemporary ethnography, yet the question of how ethnography relates to, and participates in, wider social conditions remains important. In this sense, contemporary ethnography's encounter with digital data is but a recent unfolding of the longstanding relationship between methods and the social relations they simultaneously examine and create.

These histories are far more complicated than we can address here, but one lesson we take from them is that the development of methods require a critical awareness of, and engagement with, other participants likely to use them. We believe not that scholars should avoid coming up with new methods, lest they participate in a social world one would not have wanted, but that developing new methods requires broader engagement. Indeed, anthropology has been coming to terms with the social life of its methods for quite a long time, whether in

terms of the representational politics it participates in (Clifford and Marcus 1986) or in terms of its response to the use of ethnography in other disciplines (Ingold 2014; Madsbjerg 2014).

Whilst it is possible then to construct a history of both ethnography and of the use of computing in anthropology, these brief reflections demonstrate that choices over research method are specific and contingent to the circumstances in which they take place. Even the most exhaustive history of these prior practices would therefore be insufficient to explain the current interest in digital data analysis both outside and within anthropology. To do this we must turn to the current configuration which the chapters of this book elaborate on that combines both the production and the use of new digital data sources, and the form that ethnographic practice takes within anthropology today.

Digital anthropology

For the past two decades, the main debates about computers within anthropology have come under the umbrella of what is now known now as 'digital anthropology'. The aim of digital anthropology has been to study the significance of digital phenomena, which serve both as object of ethnographic enquiry – what happens in online communities, or in data-mediated interactions – and as a methodological puzzle about how come to understand those social worlds. A rich literature on digital methods for qualitative researchers in turn has ensued (Hine 2015; Pink et al. 2016).

A key issue at the heart of digital anthropology has been the way in which 'the digital' raises theoretical questions about how reality is constituted. Notions of 'the digital' perpetuate (unfairly, Boellstorff argues) tropes of 'virtuality' and 'unreality' at the very moment that anthropologists are asking questions about how to move beyond the virtualising concept of culture. That is, in order to understand the variety of existing lifeworlds in their own terms, this persistent trope of digital cultural formations as somehow less real than other cultural forms must be rethought (Boellstorff 2016). The question of how reality is constituted has drawn digital anthropology into the heart of contemporary anthropological debates about ontology and the constitution of sameness and difference (Boellstorff 2016; Knox and Walford 2016). 'The digital' is more than just a new terrain to

interrogate, or a new set of methodological problems and opportunities, but takes us to the heart of the issue of what it looks like to take seriously other people's ontologies, and the grounds on which we can say others' worlds are 'the same' as or 'different' from our own.

One of the key anthropological critiques that has been made of data is that it is an abstraction (Carrier and Miller 1998). As a partial storyteller that strips away much of the richness of social interaction in order to render things amenable to mathematics, numbers are thus seen to form their own virtual reality (Miller 2002). Boellstorff (2012) argues in contrast that, whether an online game, a cultural construct or a physical artefact, we need not decide in advance whether something is a virtuality or a reality. We could instead see these entities as both 'real' and 'virtual' at the same time. This is helpful for rethinking how we might approach digital data ethnographically. Instead of starting from an assumption that says data's primary status is representational, the chapters in this book examine different data types as things that have both representational strategies and ontological properties. Sensor data, for example, attempts to point towards a bodily phenomenon like heart rate at some distance from the sensor technology itself, while click data is about as close as one can get to the click itself. Representations and ontologies work differently in both cases. While both can be overinterpreted, or elude the real object of study, to figure them both as primarily abstractions – socially performative abstractions perhaps but abstractions none the less – deflects attention from their reality-producing effects.

This then means we have a particularly thorny 'social life of methods' issue: ethnographers are required to take seriously the ontological status of numbers and their relationship with an underlying reality, while also taking seriously the contradictory emic injunction to always take numbers with a grain of salt, and treat them as virtual simulacra of the lifeworlds to which they refer. We must do both of these while acknowledging data's palpable materialities that also somehow shape-shift into various material forms (graphs, sounds, sensations etc.) (Berson 2015). These contradictory injunctions make for a tall order. Digital data, whether a computation of click patterns or readings from an instrumented environment, involves material

frameworks raised by Boellstorff's theorisation of the digital. As the chapters to follow show, computational forms have their own particular ways of creating and erasing both difference and sameness, and scales 'large' and 'small'.

Whose methods?

Recognising the role of methods in constituting social worlds also suggests a further possibility, namely that methods for knowing the world are problems that go beyond professional scholarship. The 'social life of methods' approach suggests that scholarly knowledge production is connected to the world 'out there', and that other people who are not scholars are also capable of creating and using methods. Indeed, when it comes to digitally collected data, techniques for knowing the social through 'transactional' data – that is, data that occurs as the result of everyday exchanges like clicking or using social media, as opposed to data collected for social research – have been elaborated far more rapidly and extensively outside of scholarship (Savage and Burrows 2007). For Savage and Burrows, these developments constitute a challenge to sociological authority, putting sociologists in the uncomfortable role of methods adopter, rather than methods creator. As described by Grommé, Ruppert and Cakici (Chapter 2 below), the introduction of data science methods into European statistical institutes created some consternation, but also the need to develop new professional practices and forms of social capital for constituting proper data science for this purpose. For many new media and communication scholars wishing to understand the cultural and social worlds of social media, this question of who gets to produce knowledge with data has become quite acute. Social media companies' convoluted methods of data handling, often hidden behind claims of intellectual property, pose real challenges to those trying to understand the mechanisms by which social media feeds or online content are organised. Indeed, one methodological intervention in this field has taken the form of an American Civil Liberties Union (ACLU) lawsuit, launched on behalf of media scholars Karrie Karahalios and Christian Sandvig, to persuade the United States govern-

which to develop methods, we do not live in a social world where we can safely presume that methods necessarily originate in scholarship.

The use of data by computer scientists, social media and consumer health companies to produce knowledge about the social raises well established issues of legitimacy, expertise, dominance and access. Expanding what we mean by method to include research done in the course of everyday living, as opposed to an exclusively professional practice, introduces a much richer set of social dynamics. Consider, for example, Noortje Marres's (2015) work on experiments in green living. Marres is interested in notions of scientific experimentation as a frame with which people come to understand what ecological homes are about. Putting in compostable toilets or solar panels have become, in certain circles, tests of what is possible. These tests shape how some people with green homes relate to one another – an ethos centred more on what was learned about energy consumption or material feasibility than the cultural identities also on display. Similarly, people in the Quantified Self movement who use data to experiment with their health, either out of curiosity or out of necessity, have developed a repertoire of 'paraclinical practices' (Greenfield 2016), procedures that appropriate clinical practices of data collection, experimentation and intervention, repurposing them for radically new ends that include narrative making as well as identifying new interventions.

These examples point towards the *everydayness* of methods as empirical devices (Marres 2017), and a richer social life of methods than the one dominated by the territory-making of high social capital professionals. The everyday use of methods for the production and communication of knowledge also extend well beyond scientific empiricism. For example, numbers regularly feature in the literary and visual arts (Connor 2016; Chilver 2014) not for the purpose of constituting scientific exactness but as a method of bringing into the world experiences of cadences, visual proportions and more evocative imaginations of bigness and smallness (Tufte 1983). When Melanesians display so many shells or towers of yams for exchange, or West African traders use intricate, deliberately crafted methods of reckoning (Guyer 2004), we see readily how methods of counting and reckoning become a nuanced part of everyday lifeworlds.

Following the social life of methods, then, means locating method both in the processes by which the author assembles his or her account

and in the social worlds of the protagonists. This book seeks to broaden commonplace understandings of where method might occur, whether that method is ethnographic or computational, or both at the same time, in order to understand more deeply how digital data is becoming implicated in the social worlds that people make. It also acknowledges how various actors are differently positioned with respect to methods of digital data collection and analysis. What presents itself as a methodological question to one actor might be mere substrate for another. For example, 'everyday' experimenters often do not have many choices about how numerical data comes into their worlds, but might be able to resituate it as part of an experimentation process. Data scientists, on the other hand, have fewer opportunities for bringing data closer to situated contexts in which it lives, but often have more complex computational repertoires at their disposal. There is a mutual interdependence at stake, even though these actors might not in the end produce a shared same social world.

Numbers and narratives

The longstanding distinction between qualitative and quantitative knowledge production lies in the background to this rejoinder of digital data and ethnography. If methods have a social life, this distinction becomes not an obstacle to be overcome but a particular social arrangement that needs to be better understood. Numbers and qualities are not inherently opposite ways of seeing the world. Numbers have semiotic qualities (Guyer 2014; Verran 2012) and do more than just measure. Anthropologists have never just ignored numbers as they encounter them in fieldwork, and, while they rarely measure or calculate numerically in creating an account, they are hardly uninterested in quantities or questions of prevalence and scale. Quantitative and qualitative knowledge are not inherently separate, but the distinction between the two has been a longstanding Western cultural cleavage that has had the effect of separating them out. Data practices and ethnographic practices have thus found themselves in different spaces but now, for both material and conceptual reasons (as the chapters of this book attest) they are being brought back together in newly conceived configurations.

To understand the opening up of new connections between these methods it is useful to remember just how much of a cultural project

it was to create the association between numbers and notions of objectivity or truthfulness. Ian Hacking (1990) reports that, in the Renaissance period, people who played with dice and coin flips began noticing the intriguing regularities that would later become Gaussian probability theory. Yet their forays were largely considered suspect. Chance was considered dicey: proper, rational thinkers believed in universal natural laws that controlled all things including whether a coin would flip heads or tails at a particular moment. It took a few hundred years for cultural assumptions to form that true and false could be not binary but probabilistic, and that there were reliable mathematical regularities that meant chance could be tamed.

Numbers as a form of attestation, and a way of producing evidence more convincing than personal testimony, began to feature in scientific practice during the Enlightenment period, when merchants also began using them in trading practices in order to improve their social standing. Numbers appear to systematise because they introduce their own, internally consistent formal grammars that are difficult to break (though, as the saying goes, statistics can be considerably much worse than damn lies). That very systematicity became a form of attestation. Numbers served as the perfect language in which mercantile trade could be formalised, through techniques like double-entry bookkeeping and weights and measures standardisation, which served as a kind of testimony that was beyond the merchant himself. The history of the modern fact, as Mary Poovey (1998) puts it, is one that created the separability of measurement from measurer in order to create a form of knowledge production that could rival the moral authority of the church. This notion of separability of measurer and measurement, and the notion of probabilities as a form of objective regularities, took years to intertwine into default twentieth-century notions of 'quantification'.

This longstanding trope of quantification as a kind of machinic objectivity, or direct line into once-and-for-all universal truths, is now under strain. As digitally produced data proliferates, the means that people have for evaluating the relationship between data and what it is supposed to represent are transforming. The unmooring of signifiers from signifieds for example, is at the heart of debates about

breakdown of a separation between the thing being measured and the effects of measurement itself has become particularly prevalent, with all manner of creative and playful subversions of algorithmic suggestions now proliferating (Merry and Conley 2011).

For data analysts both in academia and in non-academic settings, huge repositories of unstructured data poses other problems. Rob Kitchen (2014) charts the adoption and sometimes rejection of big data methods across the sciences, social sciences and humanities. Kitchen points out that one problem that scholars in the natural sciences have with big data is that it creates 'bigger haystacks' in which to conduct 'fishing expeditions' that end in spurious correlations. In general, the more parameters one has, the greater the chance of spurious correlation. Some data scientists informally say that after ten parameters spurious correlations become highly likely. This scope would all too easily be achieved if one were to try to correlate all the parameters measured by an activity tracker with data retrieved from a weather data service. Kitchen observes that these conditions make it crucial to have deeper discussions across disciplines, including the natural sciences, about the importance of critical reflexivity and acknowledging the situatedness of one's own methods in sorting out which correlation is spurious and which is not. He argues that in the broader course of scientific research that largely goes unpublished – problem selection, hypothesis selection, and early trial and error – critical reflexivity indeed is as epistemologically significant as deduction, but under-acknowledged. If early exploration becomes datafied, and if exploring a problem now involves exploring a dataset, perhaps experiments can be designed on the basis of that exploration rather than assuming that data is there for the purpose of proof. In these circumstances, Kitchen argues, unspoken practices of research might be talked about more openly. Computational methods for exploring, rather than 'proving', might in fact be what is being developed in big data. It is certainly central to the suggestion of many of the chapters of this collection.

What scientists might only ever recognise as the 'exploration phase' could be one way in which ethnographers might create an opening into big data as both method and field site. Indeed, that is precisely where ethnography's epistemological strengths lie. Following the

ethnographic enterprise. As Knox, Mackenzie and My et al. demonstrate in their chapters, ethnographers are less daunted by the large haystacks that other researchers find so troubling. To do fieldwork is to live with sensorial excess. Our haystacks are always 'too large' regardless of whether the social order we are examining is as vast as global capital flows or as seemingly 'small' as village kinship practices (Strathern 2004). They are consistently large because

social anthropology has one trick up its sleeve: the deliberate attempt to generate more data than the investigator is aware of at the time of collection. Anthropologists deploy open-ended, non-linear methods of data collection which they call ethnography ... Rather than devising protocols that will purify the data in advance of analysis, the anthropologist embarks on a participatory exercise which yields materials for which analytic protocols are often devised after the fact. (Strathern 2004: 5)

Ethnography is not, however, the infinite tracing of all threads. We cut the network at a certain point (Strathern 1996), usually the point at which we believe our gathered partialities have a plausible coherence. This is the exact opposite of the epistemological premises taken by Gaussian probability theory, the theory on which the bulk of twentieth-century quantitative methods are based. In introductions to probability theory, the examples of coin flips or dice rolls are commonly used to show how to calculate odds. Instructors use the examples because they make it easy to concretely imagine a field of infinite exact flips of the same coin. The notion of a probability distribution relies on a notion of infinite 'tries' – that if I were to consistently roll the dice, and the dice were not loaded and the table was flat and there were no other conditions to affect the outcome, we could conclusively say that the odds of rolling a six are one in six. As students learning about probabilities, we might then learn that if we were to ask African Americans the same survey questions as White Americans, a conclusive result would be the result that showed the same distribution time and time again among the two groups, as if one were rolling dice. This possibility of the repeat of the roll of the dice is crucial to the production of modern facts (here, the difference between two groups). It yields the truth effect of a once and for all claim, universal across the proclaimed boundaries of the problem. Here, a true claim would

hold across all African Americans and White Americans (unless, of course, circumstances fundamentally change).

In ethnography, however, there is no roll of the dice time and time again. Our focus is often on the contingent nature of social relations, making clean and clear repeatability difficult to imagine, and shaky grounds on which to attempt an analysis. Many of us in turn hold a suspicion of universalising intellectual manoeuvres on both epistemological and ethical grounds.

The style of statistics we have just described is radically changing in the context of big data. The centuries-old dilemmas in Hacking's work about natural laws versus probabilistic knowledge is being revisited in the current debates about 'frequentist' versus 'Bayesian' approaches. The 'frequentist' side of this debate draws on the Gaussian notions of predictability and probabilities that we have just described. The Bayesians draw inspiration from Thomas Bayes, an obscure early eighteenth-century statistician largely ignored until the advent of computer science and now celebrated as a kind of founding father of big data parsing techniques. To continue with the coin toss example, a Bayesian, like an ethnographer, would be inclined to reject the premise that there might be an infinite field of coin flips, each exactly the same as the next on the grounds that that convenient abstraction does not reflect the world as it exists. Instead, Bayesians are much more inclined to think about the situation in terms of what we can say descriptively about the conditions, and how probable it is that that situation will occur. In the coin toss example, that might translate into an attempt to calculate the probability of the angle of the coin's launch, or the force used.

This manoeuvre in some ways echoes Hacking's 'determinism' of natural laws (Hacking 1990), in as much as there are substantive claims being made about the factors in play. While Bayesians resemble Hacking's 'determinism', Bayesian calculations are far less deterministic in the contemporary usage of the word than a frequentist approach. In the frequentist approach, numbers focus on regularity of outcome. They are used to purify the experiment away from prior knowledge. Frequentists start out with a hypothesis that they show to be true or not, while in Bayesian work the hypotheses themselves can be probabilistic as well as the outputs. Bayesian approaches do not attempt to say what is true or not true in an overall manner, but make explicit,

and mobilise, prior knowledge. When that prior knowledge turns out to be wrong, a Bayesian analysis would go back and run the calculation in an adjusted manner. This builds into the mathematical procedures a stronger notion that situations are *likely* to change. This is one reason computer scientists talk about 'machine learning' as a form of big data analysis – machines create these iterative loops and refinements relatively easily. The notion of learning here is less an anthropomorphism than an admission of the idea that the results depend upon prior assumptions. To claim, as early pundits did, that big data was free of *a priori* theoretical knowledge is to misunderstand entirely the situated nature not just of data collection, but of how big data statistical techniques actually work (see also Bowker 2014).⁵

In this way, Bayesian approaches to assembling data always start with a notion of partiality and contingency that we more readily take to be ethnography's bread and butter. It is a view of probability as 'orderly opinion, and that inference from data is nothing other than the revision of such opinion in the light of relevant new information' (Kortz and Johnson 2012). Bayesians then do not attempt to 'devise protocols that will purify the data in advance of analysis', as Strathern put it (2004: 5–6), but fold in 'impure' propositions on an ongoing basis. They do exactly what Strathern says ethnographers do: revisit the analytical device used after the fact, once new information is known. (Though, as John Cheney-Lippold (2011) argues, the bounds of 'learning' in Bayesian computation are necessarily much narrower than in ethnography. Changes to how social categories are encoded in the computation are much harder to make than changing the elements associated with that category.) These incremental buildups of beliefs do not require reference to infinite sampling as the standard of evidence. Nor do they point towards a goal of universal truth. Bayesians do not tame chance, they proliferate it.

It would be false to say that the current preference for Bayesian techniques in contemporary computer science has created a flood of ethnographers using them to trace webs of cultural significance. There are very few who would even be in a position to do so. It would also be false to say that Bayesians never use frequentist techniques, or fully subscribe to the epistemological implications we have just drawn. There are likely more positivists than post-positivists currently using those techniques. Bayesian approaches to computation nevertheless do something epistemologically significant for ethnographers. They

leave open the possibility of situated knowledge production, cutaneous with narrative, in ways that frequentist approaches foreclose or render invisible.⁶

Digital data and the sociology of knowledge

With the development of new computational techniques for approaching numerical or digital data, a number of social scientists working primarily in the field of science and technology studies and the sociology of knowledge have begun to interrogate the potentially profound epistemological implications of big data analytics. Muniesa (2014), for example, has focused on financial markets to show how the semiotic aspects of a calculation – how a number is meant to mean – matters profoundly to the relationships it performs. Thrift (2004), on the other hand, has noted that the sheer quantity of calculations required in big data analytics also becomes a change in quality:

In recent years the activity of calculation has become so ubiquitous that it has entered a new phase, which I call 'qualculation' ... where many millions of calculations are continually made in the background of any encounter. It is no longer possible to think of calculation as necessarily precise. (Thrift 2004: 584)

In other work, Thrift argues that specific numerical compositions become 'calculative infrastructures' that bring a set of actors and actions into relation to one another. They create the conditions of possibility for a particular set of connections and disconnections. Often for indicators in the commonsense understanding of the term – a Consumer Price Index, or Gross National Happiness measure – their capacities to configure social relations become apparent when we ask what exactly these things indicate, and how. Who is implicated and who is excluded? Blok and Pedersen (2014) echo Thrift's 'qualculation' in their description of their own field research as a 'qualquant experiment', a concept that is extended in their chapter in this collection on transversal collaborations. They note that new polyphonies can emerge in intersections between different modalities of 'data' both ethnographic and numerical. Additionally, a recent ethnographic experiment called 'ethnomining' collected digital data deliberately for research purposes and then discussed it with the participants themselves (Anderson et al. 2009).

Latour (2002) and Latour et al. (2012) push this thinking even further, by linking the epistemological promise of big data analytics to a much longer-standing, if only recently resurrected, debate between Gabriel Tarde and Emile Durkheim. Latour et al. (2012) argue that the prospect of being able to capture, mine and newly represent digital traces represents a way of comprehending social patterns that profoundly breaks from the Durkheimian traditions that dominated twentieth-century ways of knowing the social. Durkheim saw social structuring as something that takes place outside the individual, and thus surveys were used precisely to de-individualise matters, and to decouple the hypothesis from any particular instance. In this way, surveys evoked the infinite flipping of a coin, and were largely processed in frequentist ways. Tarde, however, was sceptical of Durkheimian structure, and saw instead a perpetually emergent flow of discrete interactions between individual persons. Individual persons are not discrete, autonomous, rationally calculating individuals in the Tardian view; they are *a priori* social, configured by transactions experienced previously. The conditions of possibility created in one transaction potentiate what is conceivable in another, not unlike a Bourdieuan notion of disposition (1977). In this way, Tarde does not rely on unseen overall structure to account for coherence and patterning. Venturini and Latour see his work as questioning sociology's 'fictive distinction between micro-interactions and macro-structures' (Venturini and Latour 2010: 4) and providing a theoretical basis for the kind of rejoinder between the qualitative and the quantitative that we discussed earlier.

Venturini and Latour point out that digitisation further exposes the fictive nature of this distinction. It has also meant that a Tardian sociology is more conceivable now than it would have been in his own time, when those transactions barely left a trace. Now many of them do, if partially and inconsistently. Tarde imagined that social transactions would be countable, and even speculated about a future 'glorimeter' that would measure reputation and would invite people to make use of those metrics in ways not unlike contemporary social media. One implication, then, is that digital data created in the course of social interactions are not just 'bigger' than surveys. They are ontologically quite different, much more like the objects of material culture research or archival work. They also have the potential to

elaborate forms of knowledge that conceptualise social regularities quite differently from Durkheimian traditions.

All of these works meld 'data as thing in the world' and 'data as method' (Nafus 2016) into slightly different configurations, and tighten the connection between the two in different ways. Collectively, they make a solid case for an ontological and epistemological entanglement in a data-saturated world. This in turn raises two questions for ethnography. On the first hand, there is the question of what the appearance of big data is doing to the established opposition between quantitative and qualitative methods. Our exploration of Bayesian techniques and Strathern's partial connections has hinted at some of the reasons why ethnography and data analytics might have more in common than it seems at first sight, though the chapters to come have much more to say on the matter. On the second hand, if new forms of data are redistributing expertise, then the usual object of ethnographic critique (the reductionist quantitative researcher) and the usual response to this critique (the need to fill in the gaps with ethnography) both need revisiting. The book tries to address both. It asks how ethnography figures as a specific knowledge-producing practice in a data-saturated world, and it asks where more productive objects of critique might lie if method cannot be located *a priori* in particular researchy figures or practices. It does so by approaching data and ethnography not as things to be compared but in terms of the shared concepts and practices that appear to be at stake in current attempts at knowing the world. If knowledge production can be thought of as a kind of interdisciplinary trading zone (Galison 1996), the chapters collected here suggest that with the advent of digital data there are renewed potential grounds for tradability. At the same time, just as Hacking noted that early dice rollers had no idea they were experimenting with long-term epistemological transformations, so too does our present condition make it impossible to fully say what kind of epistemological transformation we might actually be in.

Whilst we cannot provide an ultimate answer to the relationship between digital data and its epistemological implications, this collection none the less aims to provide an orientation device that might aid our capacity to see where things are changing, where conventions are being disrupted and where forms of knowledge and social organisation look likely to endure. The way we do this is to return to our

central interest in what digital data is doing to different forms of expertise – both quantitative and qualitative.

The book is structured so as to help the reader navigate through this sometimes disorienting terrain of border crossings, combinations and experimental collaborations. It starts with a consideration of data science, understood from the perspective of ethnography. Data scientists are frequently heralded as the powerful experts of a new digital social order. The first question we pose is *who* are these experts and where did they come from? Part I addresses this question through three comparative ethnographies of emerging data science. In Chapter 2 Francisca Grommé, Evelyn Ruppert and Baki Cakici take us into the world of national statisticians in Europe to explore how statistical knowledge is being reworked in response to the appearance of new kinds of data. Focusing on the procedures by which data scientists are recruited by national statistical agencies, we learn about the core qualities and competencies that are attributed to an ideal data scientist. Here Grommé et al. trace the fault lines between established and new forms of expertise, and highlight how what is required of data scientists is not just data analysis skills but also an aptitude for collaboration, a potential for learning and a capacity to imbibe the *habitus* of what it means to be an official statistician.

In Chapter 3 Ian Lowrie takes us to Russia and to the university courses where data science is taught. Lowrie provides us with insights into the relationship between that expertise which is necessary to being a data scientist and that which differentiates data science from other disciplines. Here, we find a similar tension being played out between mathematics and data science that we saw in the previous chapter, cast here as a tension between pure and applied knowledge. In the Russian setting, to be a data scientist is to be able to turn the science of mathematics and statistics to productive and useful ends. Although hard skills constitute the heart of data science in the classroom, purely classroom-based or textbook-based instructional learning misses a huge part of what it means to be a data analyst – an observation picked up later by Joseph Dumit in the interview that concludes this book. In Russia, the urge to ‘keep learning’ is treated by Lowrie less as a process of cultivating unique skills and more a process of developing a new kind of practical craft.

In Chapter 4, Kaion Williams takes us to Jamaica where there have been ongoing attempts since the early 2000s to generate a successful

digital economy on the island. Williams’s focus on a developing economy opens up attention to the often invisible geopolitics of data analysis, charting how those who service the data analysis needs of call centres, offshore banking services and low-wage data processing have to grapple with the racialised imaginaries of what data science is and what it should be. Far from data science being a utopian or flat space of knowledge production, the chapter reminds us how fields of expertise work to reconstitute geographies of inclusion and exclusion. Here, we also get our first glimpse into the way in which ethnographic experiments with data traces might be deployed reflexively to shed light on where boundaries can and cannot be traversed. Williams introduces our first experiment in the form of an attempt to critically deploy data as a way of pushing back at the inevitable social positioning of the researcher that take place in all ethnographic research.

Part II of the book moves us from ethnographies of data science to three chapters which pay close attention to different social lives of data and the ways of knowing that they entail. Turning from the culture of data science to the ontologies of data and their entanglements with other forms of expertise including ethnography, these chapters unpack what kind of relational assumptions are at play in different data practices – both qualitative and quantitative.

In Chapter 5, Antonia Walford explores the role that databases and practices of collection play in the creation and stabilisation of environmental data. Moving beyond the idea of the database as a technology of storage, Walford draws our attention to the centrality of sharing, circulation and reuse of data within archival practice and thus expands our understanding of the relational promise that data practices enable and extend. Describing the work of those who have been trying to develop audacious new methods of global environmental data collection, Walford unravels the relational commitments of people, environments and technologies that the practice of collection and the ordering of scientific knowledge entail, highlighting how data does not do away with but rather profoundly changes the structured relationship assumed between data parts and archival wholes.

Staying with climate scientists, in Chapter 6 Hannah Knox takes up a question posed in Walford’s conclusion about the implications of the relational understanding of data for the practice of ethnography. Knox responds to this by asking whether numerical data might

provide a hinge point around which climate modellers and ethnographers could enter into a more productive dialogue. Launching from the social/natural boundary-crossing posed by the anthropocene, and focusing comparatively on how climate modelling and ethnography produce knowledge via data, Knox asks whether there might be a way of reformulating ethnography's relationship to data in a way that allows ethnographers to hold on to the reality that climate data narates, whilst still acknowledging the relational and emergent character of anthropogenic climate change. In her conclusion Knox makes a tentative proposition as to what this might look like by engaging with practices emerging from the fields of architecture, art and design.

In Chapter 7, Adrian Mackenzie asks a similar question of a very different context by posing the question: how can we do an ethnography of a very large number? In the context of big data, and of claims that are made as to the power of large numbers, an attention to the question of how we and others actually engage big numbers is crucial. Attempting to engage ethnographically with the number of code repositories stored on the GitHub platform (29 million), Mackenzie explores what a big number like this actually is, asking: How is it composed? Calculatively ordered? Used to figure specificity and difference? Large numbers are crucially shown to be both integral to the functioning of data infrastructures *and* central to the imaginaries that such infrastructures produce and proliferate. Numbers are thus shown to be not only tools that can be used by researchers to illuminate relations but also lively subjects with whom ethnographers should learn to engage in all their ontological complexity.

With a better grasp of how both data scientists and ethnographers can and do produce knowledge through data, Part III presents specific examples of how this reformulated understanding of both data analysis and ethnography is producing collaborative experiments in 'data ethnography'. In Chapter 8, Mette My Madsen, Anders Blok and Morten Axel Pedersen describe a collaborative experiment with data scientists at the University of Copenhagen which has attempted to create a picture of the 'social fabric' of a cohort of students by combining qualitative and quantitative data on their movements and practices. Madsen et al. use the term 'transversal collaboration' to denote the messy, emergent and uneven nature of the collaborative work required to produce provisionally stabilised knowledge about a shifting social field. In this description, the subject and object of research are

constantly under a process of reconfiguration, with Madsen et al. deploying ethnography simultaneously as a means of reflecting on data collection and a way of doing data collection itself.

In Chapter 9 Alison Powell reports on another experiment in which she devised a method for engaging data that she calls 'data workshops'. This method works to draw out the phenomenological dimensions of data by eliciting embodied and emotive responses to the city as a datafied landscape. By attending to how 'data workshop' participants came to engage with data as part of the fabric of urban life, Powell develops an approach to quantification that draws on ethnography, critical mapping and art practice to deepen our appreciation of what numbers *in situ* do to social relations. This method is used in this chapter to develop a critique of a functionalist vision of smart cities and their social effects, but it also presents a method which might be deployed in a much more diverse range of settings.

In Chapter 10 Dawn Nafus also describes her own experience of engaging publics with data, in this case focusing on people's relationship to increasingly ubiquitous streams of sensory data. Nafus outlines several of the lessons she has learned over the past ten years through her involvement in projects that have used qualitative, quantitative, textual and visual methods to help people think about different forms of sensory data that their devices are producing. As in Madsen et al.'s chapter, what emerges is a description both of how data science changes ethnography and of how ethnographic sensibilities affect the design of data analysis platforms. For example, when one approaches data in an ethnographic mode, flaws, gaps and ambiguities in data become not problems to be solved but sites of productive engagement through which people begin to engage, think and interrogate the relations that might lie behind data traces. That ethnography might demonstrate the qualitative potential of data itself is indeed one of the key insights of this volume and one that is taken up in our final section of the book.

The final contribution is an interview that Dawn Nafus conducted with Joseph Dumit at the University of California Davis about a new data-science curriculum that Dumit has been developing with colleagues in computer science. Drawing on his experience of speaking to Silicon Valley companies in putting together the course, Dumit characterises the kind of approach to understanding data that the course advocates as what he calls a 'third thing', that requires not just

quantitative or qualitative expertise but something that lies at the interface of these two disciplinary traditions. This third thing is indeed the thing that our collection of chapters attempts to interrogate, something which we label variously as 'data studies', 'critical social data science' or what we have called in this introduction 'data-ethnography'. Whatever we term it, our aim has been to gather together ethnographer-practitioners who are attempting to dwell in this third thing in order to offer both a theoretical and a practical orientation as to how a fresh relationship might be forged between ethnographers inspired by the world-making possibilities of digital data, and data scientists committed to using data to interrogate and understand a world of unfolding social complexity.

Notes

- 1 For example the Royal Society published a report into Machine Learning in April 2017 that outlined the challenges and potential of machine learning for contemporary society (Royal Society 2017).
- 2 This claim builds on discussions in anthropology about anthropological theory as discussed for example in the journal *Hau: Journal of Ethnographic Theory*, and in publications such as Lebrer (2017).
- 3 In this book, when we say 'data' we are generally referring to digitally collected numbers, words, images etc. We recognise of course that data can also refer to ethnographic evidence of various kinds. Different anthropologists have different views on the matter of whether the written reflections, notes, materials and photographs that mark traditional fieldwork can be called 'data'.
- 4 This is also pointed out by Nick Seaver (2014a) in a blog post on the Savage Minds website: <https://savage minds.org/2014/05/19/computers-and-sociocultural-anthropology/>.
- 5 Debates about how much substantive prior knowledge one needs in fact do rage within communities of machine learning specialists. Computer vision developers, for example, debate about whether it is better to make assumptions about the physics of light in getting a machine to discern that a picture of a dog is indeed a dog, or whether to do it on a 'purely statistical' basis. 'Purely statistical', however, does rely on some prior input – it is more that the developer has less of a notion of what that input means semantically, if anything (Suzanne Thomas, pers. comm.).
- 6 Of course, here we acknowledge that there is no shortage of frequentists who, with great care and self-reflexive sensitivity, take pains to situate what they learn through frequencies and avoid God trickery. Our point

is merely that the philosophical premises, grounded in a certain history, start with a notion of 'infinite' coin flips or dice rolls or chances to count a population in a particular way, that invites the sorts of universalising claims that ethnographers are so quick to object to

References

- Anderson, Chris. 2008. 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete'. *Wired*, 23 June. http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory [accessed 4 March 2018].
- anderson, Ken. Nafus, Dawn, Rattenbury, Tye and Aipperspach, Ryan. 2009. 'Numbers Have Qualities Too: Experiences with Ethno-Mining'. *Ethnographic Praxis in Industry Conference Proceedings* 1: 123–40.
- Bateson, Gregory. 1972. *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology*. St Albans, Australia: Paladin.
- Berson, J. 2015. *Computable Bodies: Instrumented Life and the Human Somatic Nidie*. London: Bloomsbury.
- Benjamin, Walter. 2008. *The Work of Art in the Age of Mechanical Reproduction*. London: Penguin (original German edition 1939).
- Blok, Anders and Pedersen, Morten Axel. 2014. 'Complementary Social Science? Qualitative Experiments in a Big Data World'. *Big Data & Society* 1(2): 1–6.
- Bollstorf, Tom. 2012. "'Rethinking "Digital" Anthropology'. In *Digital Anthropology*. Edited by Heather Horst and Daniel Miller. London: Bloomsbury Publishing, 39–60.
- Bollstorf, Tom. 2016. 'For Whom the Ontology Turns: Theorizing the Digital Real'. *Current Anthropology* 57(4): 387–407.
- Bollstorf, Tom and Mauret, Bill. 2015. *Data: Now Bigger and Better!*. Chicago: Prickly Paradigm Press.
- Bourdieu, P. 1977. *Outline of a Theory of Practice*. Cambridge: Cambridge University Press.
- Bowker, Geoff. 2014. 'The Data/Theory Thing'. *International Journal of Communication* 8(2043): 1795–9.
- Boyd, Danah and Crawford, Kate. 2012. 'Critical Questions for Big Data'. *Information, Communication & Society* 15(5): 662–79.
- Boyer, Dominic. 2013. *The Life Informatic: News-making in the Digital Era*. Ithaca, NY: Cornell University Press.
- Carrier, James G. and Miller, Daniel. 1998. *Virtualism: A New Political Economy*. Oxford and New York: Berg.
- Cheney-Lippold, John. 2011. 'New Algorithmic Identity: Soft Biopolitics and the Modulation of Control'. *Theory, Culture and Society* 28(6): 164–81.

- Chilver, John. 2014. 'Number Intersecting with Art: Three Studies'. *Distinktion: Scandinavian Journal of Social Theory* 15(2): 239–49.
- Clifford, James and Marcus, George E. 1986. *Writing Culture: The Poetics and Politics of Ethnography*. Berkeley, CA: University of California Press.
- Connor, Steven. 2016. *Living by Numbers: In Defence of Quantity*. London: Reaktion Books Limited.
- Dunnit, Joseph. 2004. *Picturing Personhood: Brain Scans and Biomedical Identity*. Princeton, NJ: Princeton University Press.
- Ellen, Roy. 2014. 'A Short History of Anthropology at the University of Kent, 1965–2014'. Available at: www.kent.ac.uk/sac/files/history.pdf [accessed 4 March 2018].
- Gallison, Peter. 1996. 'Computer Simulations and the Trading Zone 4'. In *The Disunity of Science: Boundaries, Contexts, and Power*. Edited by Peter Louis Gallison and David J. Stump. Stanford: Stanford University Press, 118–57.
- Geoghegan, Bernard Dionysius. 2011. 'From Information Theory to French Theory: Jacobson, Lévi-Strauss and the Cybernetic Apparatus'. *Critical Inquiry* (autumn): 96–126.
- Greenfield, Dana. 2016. 'Deep Data: Notes on the N of 1'. In *Quantified*. Edited by Dawn Nafus. Cambridge, MA: MIT Press, 123–46.
- Guyer, Jane I. 2004. *Marginal Gains: Monetary Transactions in Atlantic Africa*. Chicago: University of Chicago Press.
- Guyer, Jane. 2014. 'Percentages and Perchance: Archaic Forms in the Twenty-First Century'. *Distinktion* 15(2): 155–73.
- Hacking, Ian. 1990. *The Taming of Chance*. Cambridge: Cambridge University Press.
- Hine, Christine. 2015. *Ethnography for the Internet: Embedded, Embodied and Everyday*. London: Bloomsbury Academic.
- Hymes, Dell. 1965. *The Use of Computers in Anthropology*. London: Mouton and Co.
- Ingold, Tim. 2014. 'That's Enough about Ethnography!' *Hau: Journal of Ethnographic Theory* 4(1). <http://dx.doi.org/10.14318/hau4.1.021> [accessed 4 March 2018].
- Kirchlin, R. 2014. 'Big Data, New Epistemologies and Paradigm Shifts'. *Big Data & Society* 1(1): 1–12.
- Knox, Hannah and Walford, Antonia. 2016. 'Is There an Ontology to the Digital? Cultural Anthropology Website, Theorising the Contemporary, 24 March. <https://culanth.org/fieldsights/820-digital-ontology>.
- Kohn, Eduardo. 2013. *How Forests Think: Toward an Anthropology beyond the Human*. Berkeley: University of California Press.
- Kotz, S. and Johnson, N.L. (eds). (2012). *Breakthroughs in Statistics: Foundations and Basic Theory*. New York: Springer Science & Business Media.

- Latour, B., Jensen, Pablo, Venturini, Tommaso, (Grauwin, Sebastian and Boullier, Dominique. 2012. 'The Whole Is Always Smaller than Its Parts: A Digital Test of Gabriel Tarde's Monads'. *British Journal of Sociology* 63(4): 590–615.
- Latour, Bruno. 2002. 'Gabriel Tarde and the End of the Social'. In *The Social in Question. New Bearings in History and the Social Sciences*. Edited by Patrick Joyce. London: Routledge, 117–32.
- Lebner, Ashley. 2017. *Redescribing Relations*. London: Berghahn Books.
- Lury, Celia and Wakeford, Nina. 2012. *Inventive Methods: The Happening of the Social*. London: Routledge.
- Madsbjerg, Christian. 2014. 'Happy Birthday, Now Grow Up'. www.epicpeople.org/happy-birthday-now-grow-up [accessed 7 January 2017].
- Marres, Noortje. 2015. *Material Participation: Technology, the Environment and Everyday Publics*. Basingstoke: Palgrave Macmillan.
- Marres, Noortje. 2017. *Digital Sociology: The Reinvention of Social Research*. Cambridge: Polity Press.
- Marres, Noortje and Welevrede, Esther. 2013. 'Scrapping the Social? Issues in Real-Time Social Research'. *Journal of Cultural Economy* 6(3): 315–35.
- Maurer, Bill. 2012. 'Late to the Party: Debt and Data'. *Social Anthropology* 20: 474–81.
- Mead, Margaret. 1968. 'Cybernetics of Cybernetics'. In *Purposive Systems*. Edited by H. von Foerster, J.D. White, L.J. Peterson and J.K. Russell. New York: Spartan Books, 1–11. <http://cepa.info/2634> [accessed 4 March 2018].
- Merry, S.E. and Conley, J.M.. 2011. 'Measuring the World: Indicators, Human Rights, and Global Governance'. *Current Anthropology* 52(3): S83–S95.
- Miller, Daniel. 2002. 'Turning Callon the Right Way Up'. *Economy and Society* 31(2): 218–33.
- Muntesa, Fabian. 2014. *The Provoked Economy: Economic Reality and the Performative Turn*. London and New York: Routledge.
- Myers, Natasha. 2015. *Rendering Life Molecular: Models, Modelers, and Excitable Matter*. Durham, NC: Duke University Press.
- Nafus, Dawn. 2016. 'The Domestication of Data: Why Embracing Digital Data Means Embracing Bigger Questions'. *Ethnographic Praxis in Industry Conference Proceedings* 1: 384–99.
- Pink, Sarah, Horst, Heather, Postill, John, Hjorth, Larissa, Lewis, Tania, and Taachi, Jo. 2016. *Digital Ethnography: Principles and Practice*. Los Angeles: Sage.
- Poovey, Mary. 1998. *A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society*. Chicago and London: University of Chicago Press.

- Rappaport, Roy A. 1977. *Pigs for the Ancestors: Ritual in the Ecology of a New Guinea People*. New Haven: Yale University Press.
- The Royal Society. 2017. *Machine Learning: The Power and Promise of Computers that Learn by Example*. London: The Royal Society.
- Ruppert, Evelyn, Law, John and Savage, Mike. 2013. 'Reassembling Social Science Methods. The Challenge of Digital Devices'. *Theory Culture & Society* 30(4): 22–46.
- Savage, M. and Burrows, R. 2007. 'The Coming Crisis of Empirical Sociology'. *Sociology* 41(5): 885–99.
- Seaver, Nick. 2014a. 'Computers and Socio-Cultural Anthropology'. Blog post on <https://savageminds.org/2014/05/19/computers-and-sociocultural-anthropology/> [accessed 31 May 2017].
- Seaver, Nick. 2014b. 'Structuralism: Thinking with Computers'. Blog post on <https://savageminds.org/2014/05/21/structuralism-thinking-with-computers/> [accessed 1 June 2017].
- Seaver, Nick. 2015. 'Computing Taste: The Making of Algorithmic Music Recommendation'. Doctoral Thesis, *University of California at Irvine*.
- Shrinivasan, Ramesh, Boast, Robin, Furner, Jonathan and Becvar, Katherine M. 2009. 'Digital Museums and Diverse Cultural Knowledge: Moving Past the Traditional Catalog'. *The Information Society* 25 (4): 265–78.
- Strathern, Marilyn. 1996. 'Cutting the Network'. *Journal of the Royal Anthropological Institute* 2: 517–35.
- Strathern, Marilyn. 2000. *Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy*. London: Routledge.
- Strathern, Marilyn. 2004. *Commons and Borderlands: Working Papers on Interdisciplinarity, Accountability and the Flow of Knowledge*. Oxford: Sean Kingston Publishing.
- Suchman, Lucy, Bloomberg, Jeanette, Orr, Julian E. and Trigg, Randall. 1999. 'Reconstructing Technologies as Social Practice'. *American Behavioural Scientist* 43(3): 392–408.
- Thrift, Nigel. 2004. 'Movement-Space: The Changing Domain of Thinking Resulting from the Development of New Kinds of Spatial Awareness'. *Economy & Society* 33(4): 582–604.
- Trigg, R.H., Bodker, S. and Grönbak, K. 1991. 'Open-Ended Interaction in Cooperative Prototyping: A Video-Based Analysis'. *Scandinavian Journal of Information Systems* 3: 63–86.
- Tufte, Edward. 1983. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.
- Venturini, Tommaso and Latour, Bruno. 2010. 'The Social Fabric: Digital Traces and Quali-Quantitative Methods'. *Proceedings of Future En Seine*

- 2009: Cap Digital. www.medialab.sciences-po.fr/publications/Venturini-Latour-The_Social_Fabric.pdf [last accessed 21 February 2018].
- Verran, Helen. 2012. 'Number'. In *Inventive Methods: The Happening of the Social*. Edited by Celia Lury and Nina Wakeford. London: Routledge, 110–24.
- Verran, Helen and Christie, Michael. 2014. 'Postcolonial Databasing? Subverting Old Appropriations, Developing New Associations'. In *Subversion, Conversion, Development: Cross Cultural Knowledge Exchange and the Politics of Design*. Edited by James Leach and Lee Wilson. Cambridge, MA: MIT Press, 57–77.