

Accounting for the social: Investigating commensuration and Big Data practices at Facebook

Big Data & Society
January–June 2016: 1–16
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/2053951716631365
bds.sagepub.com



Fernando N van der Vlist

Abstract

This study explores Big Data practices at Facebook through an investigation of the role of *commensuration* or ‘the transformation of different qualities into a common metric’ in the structuration of analysis and interaction with a major online social media platform. It proposes a conceptual framework and demonstrates the empirical potential of a pragmatic approach based on reading published materials and available documentation. Facebook’s Data Warehousing and Analytics Infrastructure serves as an illustrative example to begin tracing out and describe data assemblages in more detail. In being attentive to the motivations, drivers and challenges engineers face when dealing with Big Data, it is argued that their solutions can enable and support but also constrain specific analytical and transactional capabilities or data flows between various devices and actors. The analysis thus moves beyond methodological critiques of the utility of Big Data that lack empirical support and specificity. It is further argued that analytics not just describe but also actively participate in the enactment of social worlds, thereby opening possibilities for new markets or market segments to arise. Online sociality accounts for a model of the social that makes it visible and measurable *qua* markets inviting data recontextualisation and the creation of value along multiple axes. Contra Facebook’s claim to make the web more ‘social’, an investigation of commensuration brings to the fore the question how the social is accounted for in the first place.

Keywords

Commensuration, Big Data, cultural techniques, data analytics, data infrastructure, Facebook

Introduction

There is a growing public and academic interest in ‘Big Data’ as they give rise to new ways of making sense of, doing work in, managing, or imposing control upon different aspects of the social world. Over the recent years, these developments have been welcomed by those setting out passionately the case for Big Data, open data and data infrastructures – especially in the realms of commerce and business activities, but also by governments, archives and academic research – and have been critically contested by scholars in an attempt to spark conversations about their issues and negative consequences (e.g., boyd and Crawford, 2011, 2012; Crawford and Schultz, 2014; Richards and King, 2013, 2014). These profound developments have been linked to debates around ‘the coming crisis of empirical sociology’ by Burrows and Savage (2014) and Mike Savage and Roger Burrows (2007) who focus mainly

on methodological challenges of Big Data within the social sciences’ methodological repertoire. In addition, Chris Anderson (2008), former editor-in-chief of *Wired*, has made a provocative statement proclaiming ‘the end of theory’ as the ‘data deluge’ has supposedly made the scientific method obsolete, and Rob Kitchin (2014b) has more recently used the term ‘data revolution’ in his eponymous book. More recent academic considerations, however, seem to be much more cautious to avoid slipping into mere polemics or provocation. For instance, both Kitchin (2014a, 2014b) and Evelyn

Department of Media Studies, University of Amsterdam, the Netherlands

Corresponding author:

Fernando N van der Vlist, Department of Media Studies, University of Amsterdam, Turfdragerpad 9, 1012 XT Amsterdam, the Netherlands.
Email: fernando.vandervlist@gmail.com



Creative Commons NonCommercial-NoDerivs CC-BY-NC-ND: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 License (<http://www.creativecommons.org/licenses/by-nc-nd/3.0/>) which permits non-commercial use, reproduction and distribution of the work as published without adaptation or alteration, without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Ruppert (2012, 2013) have called for the need to trace out the sociotechnical arrangements or *data assemblages* – the material arrangements and practices that ‘generate conformable spaces and the possibility of qualculation’ (Callon and Law, 2005: 731) – to better understand their formation, functioning and sustenance to accompany and undergird wider conceptual, synoptic and critical analysis with detailed empirical analysis. I argue there is a need to study the implications of Big Data for society and understand how associated practices are disrupting or reconfiguring the social, industry and business relations, expertise, methods, concepts and knowledge. Big Data constitute a variety of drivers, barriers and (domain-specific) challenges for individuals and institutions (Ekbj et al., 2014), pertaining to the question of making sense of data and the ‘data revolution’ more generally,¹ or to understand emerging practices and perspectives, potential contributions, sustaining innovations or disruptions in particular domains of application such as in the fields of econometrics, operational research and the management sciences (Einav and Levin, 2014; McAfee and Brynjolfsson, 2012; Taylor et al., 2014), business intelligence (Minelli et al., 2012), social and cultural research (Manovich, 2012), or more generally ‘how we live, work, and think’ (Mayer-Schönberger and Cukier, 2013). As data mining techniques enter other domains, data analytics, machine learning, database management and their many uses in recommendation, recognition, sorting, ranking and pattern finding are reconfigured and become increasingly mundane (Mackenzie, 2015). As Kitchin has argued, there is now a critical need to engage with these matters from philosophical and conceptual points of view, as well as through detailed empirical analysis of data assemblages (2014b: 184–185), provided the potential utility and value of data, especially at today’s levels of granularity, are unprecedented.

It is no coincidence that the current iteration of this debate focusing on Big Data practices, rather than on methodological considerations, coincides with another discourse that has undergone a similar shift of focus, and which is led by some of the same British sociologists. This other humanities-infused discourse is concerned with the ‘politics of method’ (Savage, 2010; Savage and Burrows, 2007) and examines the ‘double social life of social methods’ (Law et al., 2011; Savage, 2013) – a cross-cutting approach to thinking methods not just as instruments, but as objects of study in themselves which are embedded in and shaping the social worlds they purport to describe. In other words, they seek to understand data and devices within the assemblages they form together with other kinds of actors as possessing co-constitutive agency in the enactment or materialisation of new ways of social and cultural

being, while at the same time as new forms of social and cultural inquiry (e.g., Mair et al., 2015). Still others like Sophie Day et al. (2014) have proposed to examine a particular type of assemblage or what they term ‘number ecologies’ – extending the concept of ‘ecologies of knowledge’ (Star, 1995) – approached through the numbers and numbering practices that give rise to them. For instance, Carolin Gerlitz and Celia Lury (2014) have critically examined how the performative capacities of influence measures and other ‘participative metrics of value’ are interlinked with media to enact dynamic self-evaluating assemblages. The particular approach taken by these scholars is situated within the larger field of economic sociology, a field that ‘rediscovered the economy’ (Miller, 2001: 379) with its roots in social studies of science and technology (STS) and actor–network theory. Economic sociologists enquire into the ways in which economic phenomena like markets come into being through the various agencies exercised by both technical and social actors, and the relationships of translation or intermediation these may establish in different scenarios (Callon, 1998; Callon and Muniesa, 2005; Callon et al., 2007; Fligstein, 1993; Granovetter, 1985). Among the main contributions of early studies in STS and economic sociology has been the ‘turn to technology’ (Bijker et al., 1987; MacKenzie and Wajcman, 1985; Woolgar, 1991), as well as a reconceptualisation of the affective relationship between technology and social aspects (including behaviours) as neither simply deterministic or utilitarian in their effects or impacts, nor merely embedded in and constrained by them. Rather, complex actor–networks tend to be mutually constituted through a continuous interplay of both agential and structural factors and aspects (Bijker and Law, 1992; Callon, 1986; Granovetter, 1985; Latour, 1987; Latour and Woolgar, 1986).

The objective of this article is to contribute to these ongoing debates by tracing out such a network of associations comprised of technical objects, techniques, and the operative chains they are involved in, as seen through the conceptual lens of ‘cultural techniques’ (Macho, 2003, 2008; Siegert, 2013, 2014). Who works with Big Data, its production, storage, analysis and application? What motives and challenges drive and constrain their work? What is actually done with Big Data and what other kinds of knowledge could it help produce? On the one hand, the focus is on the coordination of a range of disparate concept and methods from within a larger genealogy or *archive* of ideas – in Foucault’s sense of the term (1970, 1972), in particular from management and accounting – that are mobilised towards the purpose of *commensuration* or ‘the transformation of different qualities into a common metric’ such as a mean, price, or ratio

(Espeland and Sauder, 2007; Espeland and Stevens, 1998: 314; Sauder and Espeland, 2009). On the other hand, commensuration is simultaneously understood as a social process and accomplishment; as having a ‘social life’ (Law et al., 2011) through which it distinguishes itself ‘according to [its] domain of application’ (Foucault, 1995: 138). As will be argued, commensuration provides a practical rationality able ‘to translate thought into the domain of reality, and to establish “in the world of persons and things” spaces and devices for acting upon those entities of which they dream and scheme’ (Espeland and Sauder, 2007; Rose and Miller, 1992: 8). Concepts and methods converging in today’s managerial and accounting practices² – a more general set of concepts and methods relating to the measurement, processing, retrieval and communication of information about economic entities³ – enable tackling problems by facilitating ‘action at a distance’ (Latour, 1987; Robson, 1994). The idea that preponderant administrative practices create the things they purport to describe, an idea informed by the work of Foucault, aligns with Wendy Espeland and Mitchell Stevens’ proposition that commensuration is fundamentally relative and transforms what it measures, producing new relations as well as new entities (1998: 318, 338–339). As such, their views align well with the idea of a ‘double social life’ introduced previously in acknowledging this particular entanglement of performative relations through which cultural techniques come to differentiate themselves from each other according to their domains of application. Moreover, they argue, ‘Investigating commensuration is important because it is ubiquitous and demands vast resources, discipline, and organization. Commensuration can radically transform the world by creating new social categories and backing them with the weight of powerful institutions’ (p. 323). What does it mean to account for – or literally take into account – online sociality on a major social media platform like Facebook? What does commensuration bring to the table in this case? How do current conceptual reconfigurations involving commensuration as technique, in conjunction with new computational technologies produce new ‘measurable’ subjects (Power, 2004: 777–778) and reshape particular forms of organisation and institutional practice into powerful ‘technologies of government’ (Rose and Miller, 1992: 183)? How can we begin to understand commensuration in online social media platforms more generally as reworking boundaries between the social, cultural and economic?

The argument is organised in three parts. The first lays out the argument in general terms, positing commensuration as a cultural technique that is part of operative chains linking technological objects and social processes together in the structuration of analysis

and interaction with social media platforms, while also playing a central role in reconfiguring them. It makes a case for studying the production and circulation of a particular kind of number and its practical utility. The second part engages with these themes and techniques more concretely and pragmatically. How and where is commensuration at work in a social media platform like Facebook? Concentrating specifically on Facebook’s Data Warehousing and Analytics Infrastructure (hereinafter DWAI), it moves beyond methodological critiques of the utility of Big Data that lack empirical support and specificity. It is notable that to a company like Facebook, data and analytics are at the core of everyday operations, where the work of programmers and non-programmers, internal applications and external products converge in their reliance on the very same infrastructures, attracting many different kinds of uses and users. Reading some of Facebook’s own publications on the topic, as well as available technical documentation is a way to begin describing this unique configuration and also gives insights into the kinds of issues and challenges driving engineers to implement certain solutions over others. If Big Data can be said to constitute challenges and opportunities that often require domain-specific solutions, then the associated practices mark a political space where multiple possible solutions compete, warranting further investigation. The third part examines the relationship between Facebook’s data infrastructure and the social and economic realities it gives rise to. How to understand the role of commensuration and other calculative agencies deployed in Big Data infrastructures in the structuring of analysis and interaction with social media platforms? How is the social accounted for and what makes it that these data can become economically valuable?

Commensuration as cultural technique

Following Thomas Macho’s initial definition, ‘Cultural techniques – such as writing, reading, painting, counting, making music are always older than the concepts that are generated from them’ (2003: 179). They are conceived as operative chains. Accordingly, symbolic work requires specific cultural techniques: ‘we may talk about recipes or hunting practices, represent a fire in pictorial or dramatic form, or sketch a new building, but in order to do so we need to avail ourselves of the techniques of symbolic work, which is to say, we are not making a fire, hunting, cooking, or building at that very moment’ (2008: 100). Understanding commensuration as cultural technique means acknowledging it as an integral component in a series of actions that may give rise to symbolic and material practices. Further, it is instructive to distinguish between commensuration as

routine technique without significant consequences on the one hand, and as having technical, symbolic, or political advantages on the other (Espeland and Stevens, 1998: 316; Feldman and March, 1981). In a more concrete sense, this layering enables a meaningful distinction between the rather simple or mundane procedures and those involving tremendous collective effort. For example, distinguishing between simple routine counting procedures and shared collective counting procedures involving large infrastructures, powerful institutions and standards, or distinguishing simple acts of value comparison from high-frequency trading across cultural or geographical distances fundamental to global markets. Moreover, this perspective reminds us that these techniques are deeply cultural, historical, and open to critical scrutiny. Situating techniques within their larger conceptual spaces can enable a better understanding of the concepts and methods they mobilise.

Metrics and numbers do not only count, but also facilitate the analysis, evaluation and efficient management or control of a broad range of human activities and practices represented by these quantities. In *Control through Communication* ([1989] 1993), JoAnne Yates has traced how the diffusion of ideas revolving around what has been designated 'systematic management' by Joseph Litterer (1959) in conjunction with new communication technologies, has reconfigured internal organisational communication systems in firms into powerful systems of management and control still operative today. Recognising this history of control through communication may help illuminate present-day issues or future adaptation and innovation of record-keeping technologies in their social and cultural contexts (Yates, 1993: 275; Ketelaar, 2006: 71), and at the same time this history also contains the origins for *organisational learning* – the creation, retention and transference of knowledge within an organisation (Argote, 2013). In particular, Yates' research shows how a specific and new managerial philosophy could emerge, how it was gradually implemented in workplaces by committed managers who played a key role in introducing these new management methods and communication mechanisms, and how it produced new genres of communications (e.g., internal communications). The development of (hierarchical) internal communication systems, the management or control of their operations, and evaluation on the basis of 'flows of information and orders', 'was not simply an incidental by-product of their growth. Rather, firm growth precipitated a search for new theories and methods of management that would help achieve *efficient coordination* of large, multinational firms' (Yates, 1993: xvii, emphasis added). Current global networked communication systems have opened possibilities to

firms for still wider markets and even more scattered production facilities. But rather than simply extending existing patterns of communication, the underlying managerial and control issues provide a useful entry point to examine new operative procedures of decision-making that work on the basis of systematic capturing and analysis of Big Data. In short, cultural techniques like commensuration are interesting because they highlight particular types of work and control.

Commensuration and the work of accounting

Commensuration lies at the heart of many Big Data analytics practices, constituting a linchpin in these networks of technologies and techniques, concepts and methods converging in the form of management, control and accounting procedures. Before an analysis can be conducted on the basis of a set of qualities or quantities (e.g., observations, frequencies, or ratios) they first have to be combined or grouped together as homogeneous in order to produce a single index number. Since processes of quantification often involve some form of judgement the concept of 'qualculation' (Callon and Law, 2005) seems better suited to analyse these calculations as accomplishments that require a certain kind of work. As Espeland and Stevens explain, 'Whether it takes the form of rankings, ratios, or elusive practices, whether it is used to inform consumers and judge competitors, assuage a guilty conscience, or represent disparate forms of value, commensuration is crucial to how we categorise and make sense of the world' (1998: 314). Commensuration enables rendering certain aspects of life visible or privilege them, while rendering others invisible or irrelevant. As a general technique commensuration is often deployed to negotiate difficult contradictions (e.g., when using a mean or count to compare two or more sets of numbers), as part of routine decision-making, and as vehicle for rationalisation, while presuming that these things can be measured (i.e., assigned quantities) in the first place. At the same time, because the reasons why we commensurate can vary greatly, it is arguably important to consider the forms we use to do so, as well as take note of those who resist it for its practical and political effects. This is especially true now that commensuration increasingly participates in decision-making processes in various domains are automated (through computational algorithms) in a desire to manage uncertainty, impose control, or secure legitimacy on an unprecedented pace and scale. Commensuration exists in relation to the incommensurable, and operates by '[creating] relations between attributes or dimensions where value is revealed in the comparison' (p. 318). The relations created between various qualities thus come to constitute a common metric; a single number like 'water quality' which

arises from the aggregation of an array of other disparate attributes or dimensions such as temperature, turbidity and pH. In turn, this new common metric not only offers not just a way of *knowing* the quality of water, but can then also be *acted upon* from a distance to filter, sort, rank, order, compare and contrast different entities.

Creating and apprehending these relations of commensuration requires material and social effort, which is why it is necessary to attend to material arrangements and practices (Callon and Law, 2005: 719). Relations between symbolic objects need to be formed, sustained, kept, and verified, and the practical tasks involved in these matters typically require tremendous organisation and discipline which largely become invisible when they have degraded into everyday routines and work practices. For instance, through efforts of institutionalisation and standardisation, ‘performing some highly elaborated modes of commensuration, such as generating identical units of value in stocks or commodities futures...are complex technical feats that seem “natural” to traders and stockholders nevertheless’ (Espeland and Stevens, 1998: 318; Porter, 1995). Accounting, in this sense, provides a set of related techniques, rationales and practices for doing this kind of work efficiently: to keep and verify accounts (quite literally in the context of social media user accounts), allowing one to *reason* with them (manipulate or work with, e.g., optimise). A common way to understand quantification in accounting is therefore through the *representational accuracy* of index numbers: ‘(accounting) knowledge exists, and can be judged, on the basis of its correspondence to the external world’ (Keat and Urry, 2011: 20–22; Robson, 1994: 46). The efficacy of any managerial or accounting procedure fundamentally relies on the production of these ‘inscriptions’ (Latour, 1987; Robson, 1992) and the commensurative work they involve to reconceptualise, deindividualise and reconfigure relations to foster an ‘objective’ stance towards social processes achieved through *shared counting procedures*, which render disparate attributes or dimensions of the world – or aspects of it – formally and mechanically comparable (e.g., using counts of posts, likes, or shares). In other words, such calculative practices are effective *despite* lacking seamless representational accuracy. Not every post, Like, or share on Facebook is equal, despite the fact they can be counted and compared. It is this capacity to create and reason with precise relations between virtually anything – it ‘simultaneously overcomes distance (by creating ties between things where none before had existed) and imposes distance (by expressing value in such abstract, remote ways)’ (Espeland and Stevens, 1998: 324; Goody, 1986) – that makes commensuration into such a powerful

technique. But in order to achieve this it has to be presupposed that disparate and distinctive values – usually involving diverse forms of knowledge and preferences – can in fact be expressed in a common standardised metric in the first place, without losing information or altering meanings relevant to derivative calculations and decision-making (Espeland and Stevens, 1998: 324). As Keith Robson has argued while stressing the importance of studying techniques, ‘while accounting “knowledge” may not “correspond” to any remote context in terms of truth, this failure of reference may involve reformulating the problem of truth into a problem of power and distance’ (1992: 704). This requires shifting the focus of analysis to a more concrete level: where and how do we actually find these techniques ‘in action’ in online social media?

Engineering Big Data at Facebook

Having introduced commensuration as a more general cultural technique central to managerial procedures related to accounting,⁴ this section proceeds to further situate commensuration within its field of operation (Derrida, 1982; Foucault, 1995: 138) to examine the role it plays in constituting or sustaining an assemblage of Big Data techniques and practices. Instead of a methods-driven analysis using Facebook data, I propose a kind of empirical enquiry that focuses mainly on reading published materials and available documentation to gain insight into the motivations, problems and challenges driving and constraining the design and development of Big Data infrastructures. Here, Facebook’s DWAI will serve as an example to illustrate the workings of a data assemblage. Viewing Big Data in terms of techniques helps to see that there are, for instance, a number of specific applications that indirectly rely on processing large quantities of data such as search queries, recommendations and content filtering. In fact, the scalable analysis of large data sets is among the core functions of a number of teams at Facebook – both engineering and non-engineering – and may vary ‘from simple reporting and business intelligence applications that generate aggregated measurements across different dimensions to the more advanced machine learning applications that build models on training data’ (Thusoo et al., 2010: 1013). Facebook’s DWAI – indeed an integral component of its infrastructure (Menon, 2012) – supports such batch-oriented analytics practices, which may include such things as reporting applications like Insights for Facebook Advertisers, creation of business intelligence dashboards, or doing more advanced calculations for site features like suggesting friend recommendations to Facebook users or combining messages, chat and email into a real-time

conversation (Aiyer et al., 2012; Menon, 2012; Thusoo et al., 2010). As such, Facebook provides a rich set of tools for different kinds of users to perform analytics queries on its data.

A software engineer may perceive Big Data as posing big problems in need of working solutions that meet a certain set of criteria. While it is often difficult to obtain detailed empirical knowledge of the inner workings and management of these information systems, a general picture can be drawn based on information taken from published materials and available documentation from reports, periodicals, conference proceedings, presentation slides, blogs, technical documentation and handbooks, some of which are (co)authored by members of the teams at Facebook responsible for engineering these systems. For this reason, it also helps that Facebook's data infrastructure is built 'largely on top of open-source technologies such as Apache Hadoop, HDFS, MapReduce and Hive' (Menon, 2012: 31), for which up-to-date documentation is usually available online. Using multiple sources of documentation, it is possible to develop a provisional understanding of how these systems may work and work together, enabling data warehousing and analytics operations to facilitate day-to-day operations. Moreover, it also gives a high-level overview of Facebook's data infrastructure and enables distinguishing between some its cornerstone applications. As Aravind Menon and others have shown (e.g., Aiyer et al., 2012; Borthakur et al., 2011; Thusoo et al., 2010) there are three main components in Facebook's infrastructure. Firstly, a MySQL/DB and caching component for primary data repository. This is a relational database management system (RDBMS) based on a model increasingly challenged by current demands like iterating over billions of rows at a time or working with richly connected entities. In such cases, NoSQL or graph databases and data models developed in response to the shortcomings of relational models offer clear advantages in terms of scale and agility, which is why it is surprising to learn this RDBMS is apparently still in use. Secondly, a HDFS/MapReduce/Hive component for conducting analytics on Facebook data. Thirdly, a HBase component to run transactional applications (most of which involve data documenting user operations), which is used both for internal applications and external products (Menon, 2012: 31). These components (accounting for two types of processing, which are analytical and transactional) constitute the infrastructural groundwork for a great variety of Facebook's day-to-day operations, applications, site features and external products that involve processing large numbers of online transactions of operational data to control and manage diverse operations. The following sections first describe the cornerstones of

Facebook's Data Warehouse platform and then discuss large-scale data mining and analytics in more detail. The analysis relies on being sensitive to practical challenges (often domain-specific issues) and opportunities engineers may face when dealing with Big Data as a problem. Just as in other fields and industries, Big Data intervene and disrupt simply by posing challenges and opportunities for computer science and engineering, for example in terms of 'volume, velocity, and variety' (Beyer, 2011) – for 'Big Data is [*sic*] data that either is *too* large, grows *too* fast, or *does not fit* into traditional architectures' (Ahuja and Moore, 2013: 62, emphasis added; Krishnan, 2013) – including for companies like Facebook, Google, Twitter, and LinkedIn. Indeed, processing huge quantities of transactions – much of it 'just' moving data entities documenting user operations like posts, likes, or shares around – can involve significant financial costs and other valuable and limited resources. Each procedure and calculation takes time, requires computational resources, and electricity, the costs of which quickly rise as datasets grow. It has an economics of operation; income and revenue streams as well as operational costs need to be managed efficiently to make a model commercially viable. Additionally, with regard to analytical processing there is always a trade-off to consider between smaller datasets or sampling techniques and Big Data sets or analysis over whole populations in terms of the types of analysis it affords doing. For example, Big Data facilitates 'distant reading' (Moretti, 2013) and hypothesis-led modelling grounded in practical data relationships.

Facebook's DWAI

The concept of *data warehousing* is indicative of the increased consciousness that organisational activities can indeed be organised on the basis of data sources, and that all relevant activities can be sufficiently understood as mere transactions. The main components of Facebook's Data Warehouse platform – Scribe, HDFS/MapReduce, Hive, HiPal, and NoCron (Ahuja and Moore, 2013; Menon, 2012; Thusoo et al., 2010) – reflect these notions and the challenges associated with them, which place very strong requirements on the data processing infrastructure. In fact, each of these individual components seems to be geared (and to some extent optimised) towards addressing challenges relating primarily to *diversity* (e.g., diverse users, diverse task characteristics) and *scalability* (e.g., tremendous data growth, rapidly growing user base) that are claimed to be supported at the core of these solutions and technologies (Hu et al., 2014; Krishnan, 2013; Thusoo et al., 2010: 1013). Firstly, Scribe is mentioned, which is now an archived project repository

on GitHub no longer updated and supported by Facebook, indicating the infrastructure has undergone changes.⁵ It was responsible for aggregating log data (e.g., page actions such as likes and clicks) streamed in real-time from the web server tier and makes it available in the Hadoop Distributed File System (HDFS) cluster for subsequent analytical operations. Secondly, the HDFS/MapReduce components are mentioned and serve as the core for Facebook's 'data analytics engine'. Apache's HDFS is a popular open-source distributed file system modelled on the Google File System (GFS) designed to meet a large demand of batch processing needs (Ghemawat et al., 2003). For example, the Facebook Messages feature introduced in February 2011 – collapsing SMS, chat, email and Messages into seamless messaging, conversation history and social inbox (Hicks, 2011) – requires a level of consistency, availability, partition tolerance, data modelling and scalability apparently unmet by other systems at the time (Borthakur, 2013; Borthakur et al., 2011; Shvachko, 2010). The same is true for the more recent introduction of a new keyword search option which significantly expands Facebook's Graph Search feature, now enabling users to retrieve old News Feed posts by friends, the count of which is estimated to add up to over one trillion posts in total (Constine, 2014a, 2014b, 2014c).⁶ Another example is Facebook's ongoing DeepFace project at its AI Research lab⁷ able to automatically identify – with an astounding accuracy of 97.25 percent (Simonite, 2014) – and subsequently tag human faces using computer vision and pattern recognition techniques driven by a new approach known as 'deep learning' (e.g., Chayka, 2014; Etherington, 2014; Stone et al., 2008).⁸ Hadoop MapReduce, the other main component of Facebook's data analytics engine, is inspired by Google's map-reduce infrastructure (Dean and Ghemawat, 2008), which is a flexible data processing tool for handling (very) large data sets. The model, however, is currently challenged in favour of others like the open standard Predictive Modelling Markup Language (PMML) that are argued to be better suited for dealing with real-time applications and other limitations of Hadoop (Agneeswaran, 2014; Hu et al., 2014). Thirdly, Hive is mentioned, which is a very-large-scale data warehouse built on Apache's Hadoop platform (Thusoo et al., 2011, 2009) and is made accessible through HiveQL (HQL) – an SQL-style language for querying and performing analysis on large volumes of 'structured' data stored in Hadoop HDFS. Together, Hive and HiveQL facilitate such things as data summarising, ad-hoc queries, and other forms of basic analysis simple enough so that non-programmers are able to run analytics queries on Facebook data in other branches of the company (Menon, 2012: 31–32). This is because as

a structured query language, HiveQL enables programmers to model and query a set of practical data relationships. Yet while diversity and scalability may be supported by Hive, this is at the cost of the more expressive capacities of SQL from which its specifications originally derived. Fourthly, HiPal is mentioned, which is a data analytics tool used for distributed data analysis and has an interface for users unfamiliar with SQL-syntax to build queries on top of the Hive system. This matters greatly because data exploration and analysis are important not only to Facebook's engineering teams, but to practically all branches of the company, while at the same time not everyone is sufficiently familiar with the Hive language (Lindsay, 2009). This capability thus enables a more efficient distribution of valuable resources like expertise and skill across teams. Finally, NoCron is mentioned, which is a framework for automating repetitive tasks. It allows users, again both engineers and non-engineers, to organise their tasks into workflows with specified parameters, such as dependencies between tasks in a workflow, frequency and urgency with which the job should be run (Menon, 2012: 32).

As these descriptions indicate, what is central to this data warehousing platform is a rich set of tools for different kinds of users, enabling end users as well as internal applications, external products and third parties to perform analytics queries on Facebook data. While some of these tools may have friendly user-facing interfaces in the form of site features like Page Insights, this is not necessarily the case. Indeed, such applications are often unavailable to users of Big Data more generally; demanding technical knowledge, skills and expertise in statistics and programming now part, for instance, of the standard training for most economists (Taylor et al., 2014: 3). Furthermore, working with Big Data increasingly requires statistical techniques appropriate for dealing with entire populations of data, rather than with samples. This is indicative of the way in which analytics are moving from a *descriptive* mode (e.g., summarising samples with mathematical functions like sum, average and count to describe historical data) to a *predictive* mode (e.g., using modelling, machine learning and data mining techniques to analyse streaming or real-time and historical data to predict possible outcomes or the probability of an outcome occurring), and finally a *prescriptive* mode marked by a synthesis of Big Data and deep analytics to understand as well as actively intervene or change possible outcomes by means of suggesting decision options to take advantage of predictions (Evans and Lindner, 2012; Lustig et al., 2010).⁹ In this regard it is interesting to note that self-descriptive statements taken from a variety of user-facing predictive analytics services (as well as from presentation slides originating

in industry) largely adhere to the same patterns. Consider the following examples: ‘Measure and optimise your app’s performance on and off of Facebook’ (Facebook Platform Insights); ‘turning data insights into action’, ‘Google Analytics gives you insights you can turn into real results’, and ‘Taking action has never been easier’ (Google Analytics); ‘Twitter Card analytics gives you insight into how your content is being shared on Twitter... learn how you can improve key metrics’ (Twitter Analytics); ‘Turn data into action’, ‘Turning Big Data Into Action’, and ‘Embed Your Insights and Take Immediate Action’ (RapidMiner). In such cases, ‘turning into action’ usually requires the commensuration and subsequent analysis of different dimensions of behavioural data from users or systems to accomplish optimisation or profitable action and improve such things as efficiency or optimisation, strategic recommendations, advertising strategies, and other forms of data-driven decision-making. A procedure of commensuration is thus implemented in which analytics is deployed to leverage Big Data for actionable insights and ultimately turn insights into some kind of benefit or value.

Large-scale data mining and analytics

Data mining techniques and algorithms (e.g., predictive analytics, data analytics, pattern recognition, and machine learning) play an important role in automatic and distributed data processing and analytics across a wide spectrum of domains involving often consequential decisions about human beings (Crawford, 2013; Hardt, 2014; Govindaraju et al., n.d.). The practical requirements that predictive and prescriptive modes of analytics therefore place on data infrastructures can be very demanding. The kinds of managerial procedures traced by Yates (1993) have transformed and now exist in, for example, HDFS or Hive/HBase components comprised of both analytical and transactional processing. Similarly, the notion of organisational learning has arguably reincarnated in the form of data and information management using computer learning methods or *machine learning* – a field of study that concentrates on ‘induction algorithms and on other algorithms that can be said to “learn” [using] training examples [which] are... assumed to be supplied by a previous stage of the knowledge discovery process’ rather than ‘externally supplied’ (Kohavi and Provost, 1998: 273; Saitta and Neri, 1998).¹⁰ Although the aforementioned MapReduce framework is not a machine learning method, it is instructive to consider as it is at the core of Facebook’s data analytics engine as well as representing a more general method. MapReduce handles data sets in a two-step procedure, first splitting the input data-set into independent

chunks, which are processed by a ‘map’ function specified by a user (i.e., a query) in a completely parallel manner where many calculations are taking place simultaneously (e.g., to divide larger problems into smaller ones that may be solved in parallel). This step generates intermediary key/value pairs, which are then used by the ‘reduce’ function to link two data entities to each other by merging all intermediate values associated with the same intermediate key (Apache Software Foundation, 2013; Chu et al., 2006; Dean and Ghemawat, 2010: 72, 2008; Yang et al., 2007). In other words, the former operation performs filtering and sorting on the results of a query (‘map’) followed by some summary operation (e.g., counting the number of query matches in a search space or a URL access frequency) in the second step (‘reduce’) and thereby transforming Big Data into useful data. This analytical flexibility to draw things together is crucial to understand data as *multiple*. Instead of mere simplification or reduction, data can always be arranged in a multiplicity of variations or subjected to a myriad of analytical techniques in the pursuit of identity and differentiation (Mackenzie, 2011, 2014; Mackenzie and McNally, 2013; Ruppert, 2012). MapReduce can be seen as an attempt to retain analytical expressivity while working with Big Data in ways that go beyond what traditional approaches can handle. Here, the work of accounting and commensuration become concrete, for example by enabling users to conveniently sort indexed data in any number of ways or by offering users possibilities to specify a ‘combiner’ function, enabling partial merging of data fields to speed up operations at the cost of granularity. The other more general contribution of the MapReduce framework lies in supporting sufficient scalability (with respect to the size or volume of a data-set) and fault-tolerance for these tasks – enabling a system to continue operating in the event of a failure since all operations can be mirrored and run on two or more duplicate systems simultaneously. Both properties are important characteristics of a Big Data analytics infrastructure such as Facebook’s, where data do not just sit in databases, but are continuously managed for analytical and transactional purposes as well as for other data-driven control and decision-making practices such as recruiting and finances as well as experimentation, design and user experience, optimisation and evaluation of internal applications, services, or external products. This is also the work of accounting and of commensuration.

Like MapReduce, Support Vector Machines (SVMs) constitute a more general method. SVMs are ‘supervised’ learning models or classifier algorithms that use training data to learn to solve classification problems. They can be found at work in Facebook’s applications for face recognition (Becker and Ortiz, 2009),

identifying user behaviour patterns (Bozkır et al., 2010), or indeed for any other two-group classification problem. In this context, soft-margin SVMs are especially useful because they do relatively well with examples that are difficult to label (Cortes and Vapnik, 1995), a problem typically faced when mining social and user data as most of it is ‘unstructured’ (e.g., posts, pictures, and videos, but not likes, locations, or birthdays). Despite myriad benefits, however, there are also issues with these methods of classification, not least because of their reliance on supervision. This relates to what Solon Barocas and Andrew Selbst (2014) have termed Big Data’s ‘disparate impact’, a ‘procedural unfairness’ with regard to the complex forms of discrimination implicit in these techniques, running against common misconceptions that algorithms in general are fair or ‘neutral’, or can be made as such by ‘correcting’ for errors. Instead, fair classification is achieved ‘through a more thorough stamping out of prejudice and bias’ (Barocas and Selbst, 2014: 59), which requires tremendous effort as well as accepting that some degree of disparate impact is practically inevitable. But what amount is tolerable in a specific context? Classification thus requires compromise: fairness of specific outcomes at the expense of practical utility. Fundamental to most of today’s data mining techniques is the concept of *learning*, generally understood as taking historical data about a decision problem to produce and iteratively refine a decision rule or ‘classifier’ to be applied to future instances of that problem, and to do so automatically. Especially where such training processes are ‘supervised’, normativity and judgement inevitably intervene. In simple terms, supervised learning methods work by making inferences, or mathematically formalised leaps of faith regarding the decision how new examples should likely be labelled. In effect, machine learning ‘is not, by default, fair or just in any meaningful way’ (Hardt, 2014). Yet interference of judgement in training learning algorithms is not the only issue with these methods. There are also issues with regard to classification accuracy, typically to the detriment of minority groups. As Barocas and Selbst explain, this is due to the proportionally smaller amount of data available for those groups and the higher error rates associated with those smaller sample sizes. Consequently, the (linear) classifier learned by the algorithm is trained differently and will typically have a higher error rate for smaller samples. This is not merely a statistical issue since these methods are increasingly used to make consequential decisions about human beings. In fact, Barocas and Selbst claim, ‘the only way to ensure that decisions do not place at systematic relative disadvantage members of protected classes is to reduce the overall accuracy of all determinations’ (2014: 53). Their findings then have implications

for any similar analytical operation involving both majority and minority groups at the same time. The soft-margin method of an SVM algorithm deals with this inherent limitation by measuring the degree of misclassification so as to split examples as accurately as possible (controlled by a parameter ‘*C*’), yet still maximise the distances to the nearest examples classified with a sufficient level of certainty (i.e., good enough for current analytical purposes). In contrast to hard-margin classification methods, the classifier is thus much less sensitive to noise or outliers in datasets because it trades separability for stability, thereby making it even more problematic to generalise results. The challenges addressed here illustrate the need to study not merely users and usage, but also the producers, production and productivity of these analytical techniques. They introduce their own issues and therefore constitute a site of politics and competition.¹¹ If the object of social analytics is indeed the *socius*, or understanding ‘*commonness* across cultures’ (Schmidt, 1996, emphasis added), then commensuration is clearly an integral part of its production.

Conclusions: Accounting for economic markets?

This article set out to investigate what it means to account for – or literally take into account – the ‘social’ as it manifests on a major social media platform like Facebook; how we can understand the role of commensuration in the structuration of analysis and interaction with online social media platforms, and ultimately as reworking the boundary between the social, cultural and economic. Commensuration was conceived as a linchpin in establishing relations between technological objects and social processes involving many different practices, rationales, techniques, numbers, metrics and values; a cultural technique that may be encountered ‘in the wild’ as a basic ‘qualculative’ operation or as part of lengthy operative chains geared towards achieving a set of practical aims governing the formation, functioning and sustenance of data assemblages (e.g., optimising a recommender system). Facebook’s DWAI served as an illustrative case for describing – pragmatically and in descriptive-empirical terms – one such assemblage ‘composed of a set of apparatus and elements that are variously scaled (e.g., from local organisations and materialities to dispersed teams, national and supranational laws, and global markets) but are nonetheless bound in a unique constellation’ (Kitchin, 2014b: 186). Taking commensuration as a cultural technique involving both symbolic and material work, this article has proposed a conceptual framework for studying online social media platforms and how they relate to Big

Data more generally, and demonstrates the empirical potential of a pragmatic approach grounded in reading published documents and available materials. Is it also possible to characterise the role of these techniques and operative procedures deployed in Big Data assemblages in performing online social networked environments *qua* markets (cf. Callon, 1998)?

Extending Yates' insights that new communication technologies have opened possibilities for wider markets and more scattered production facilities to firms as well as insights from others working in the field of economic sociology, the enabling of new data flows between devices and other actors (e.g., by implementing new features and techniques) contributes to redefining existing power relations or indeed produce new ones (e.g., Gerlitz and Helmond, 2013), thereby generating or reinforcing existing power relations and inequality (e.g., Andrejevic, 2014; Barocas and Selbst, 2014; Richards and King, 2013, 2014). Those who conduct analysis on social media data can (strategically) affect markets merely by stating or visualising what they believe their users are doing, should do, or will do in the future (cf. MacKenzie, 2007). Rather than 'objective' observation, data analytics can become performative of the very phenomena it purports to describe, analyse or predict. Yet while traditional approaches in statistics were generally limited in the number of variables used (e.g., for practical reasons), contemporary computational methods are not limited in the same ways and are particularly well-suited to work on any problem using a very high number of vectors or dimensions in an analysis. Data attributes or 'features' can be selected for their usefulness or relevance to a learning algorithm for solving a specific problem. For example, signals from Facebook users can be used to perform analytics across any number of dimensions by drawing together (i.e., through an act of commensuration) any number of disparate signals to explore or 'discover' new data relationships. This not only means that data have become more useful, or that their usefulness has extended deep into other domains, but also that the *analytical flexibility* associated with data has increased quite significantly, which is to say this capacity for flexibility is dependent on the specific properties of the data infrastructure. As such, each data object from the moment it is captured and stored can entertain a multiplicity of meanings and analytical objectives for various relevant social groups. This may include human actors like users and producers of analytics, engineers, journalists and economists as well as devices, applications and systems.

In addition, machine learning also enables the 'discovery' – typically through *inference* – of additional attributes currently absent from such feature spaces, thereby introducing entirely new dimensions and categories.¹² Such possibilities thus facilitate the production

of new 'measurable' subjects as well as new markets to emerge with associated practices, rationales and techniques enabled by or directly grounded in social media data. This means that Big Data sources not only account for measured online social activity, but also enable owners and third parties to make social activities economically valuable through internal applications or external products (provided that access points are available to those data). Understanding such data flows not only requires investigating data as prefiguring its analysis, that is 'as framed and framing... according to the uses to which they are and can be put' (Gitelman, 2013: 5), but also means acknowledging their material expression through operations performed on them as part of 'calculative collective devices' (Callon and Muniesa, 2005; Callon et al., 2007).¹³ Such a perspective shifts attention to commensuration and the calculative agencies that make differences between cultures and societies visible and measurable in the capacity of markets. Social and cultural differences as expressed through social media platforms facilitate the invention and production of entirely new data-driven market segments. Data-driven markets thrive on such differences, and the representational accuracy pursued by traditional accounting numbers has become less relevant in cycles of value creation in networks and online communities. Entities produced through commensuration can be made into economic entities, which means they can become part of economic activities and of the control of economic resources. As a result, the operations performed on data entities and the new relations produced between them can affect the shaping of cultures, societies and economic markets.

As this article demonstrates, a descriptive-empirical investigation may provide useful ways to study pragmatically some of the symbolic and material mechanisms and processes by which economic entities and agents are constructed in the context of a major social media platform. In examining Facebook's DWAI, and describing its cornerstone applications as well as some major challenges, I argue it is possible to gain a better understanding of how Big Data are controlled, produced, stored, analysed and applied. The relation between managerial and accounting procedures and the social activity that numbers and metrics supposedly reflect or represent is arbitrary and involves commensurative work, which, as argued above, is both symbolic and material. Crucially, this relation is productive precisely because it is arbitrary, or as Espeland and Sauder explain: 'Numbers are easy to dislodge from local contexts and reinsert in more remote contexts. Because numbers decontextualise so thoroughly, they *invite* recontextualisation' (2007: 18, emphasis in original). Furthermore, it is noteworthy that the kind of techniques and operations deployed in Facebook's DWAI

for supporting its social networking practices are quite similar to those described by Callon and Muniesa (2005) as responsible for performing economic markets. In particular, commensuration and ‘calculative agencies’ (Callon, 1998: 3) at work in Big Data infrastructures enact specific forms of socioeconomic organisation, which like other market structures are marked by internal tendencies and regularities, statistical patterns, trends and laws of behaviour, and where socioeconomic phenomena – and feature spaces more generally – can be measured, mapped, ordered, evaluated, sorted, filtered, summarised, ranked and can have states, prices, ratios and values. Such entities need not be economic, but can obtain this quality as a function of their existence within a market structure that *governs* its relations to other entities in a distinctive way through shared counting procedures. Commensuration is central to a conception of online social networks as markets, coordinating social and cultural activities and data flows. However, whereas economic entities can be quantified, compared, traded, controlled and managed in terms of tendencies, flows, prices, values or ratios, non-economic entities do not necessarily follow the same logic and need to be *made* commensurable first.

In conclusion, I propose four directions for further enquiry. First, following the approach proposed in this article, cultural techniques may be studied at different scales, varying from their simplest forms or their implementation in extremely sophisticated operative chains like those found in Facebook’s DWAI that require their own infrastructures designed specifically to accommodate such operations. This means understanding techniques as in terms of their positions within operative procedures – the techniques and cultural formations preceding them and new social realities they give rise to – as well as understanding the role they play in coordinating technological objects and social processes within larger assemblages. When performing analytics or aggregating numbers, we do not just count, but actively participate in calculation and the enactment of social worlds. Second, extending Kitchin’s call for more critical and philosophical engagement as well as detailed empirical research on the formation, functioning and sustenance of data assemblages, I suggest to study Big Data as constituting challenges addressed differently across domains such as economics, engineering or government, each of which has its own distinct rationales and practices. This includes investigating the work of engineers dealing with Big Data not just as a source but as a challenge in need of a working solution. Third, data are never simply there, but should be understood simultaneously as abstractions and as situated material-semiotic entities because these may assemble relations differently as in ‘second-order measurement’ (Power, 2004: 771) or further

aggregations of data and numbers via statistical and mathematical operations. In particular, a sensitive attitude is needed toward the commensurative processes involved in prefiguring data, as well as the analytical operations we perform on them. The challenge is to distinguish general relational qualities of data and numbers from the specificities they gain by being situated within ‘number ecologies’ (Day et al., 2014). Investigating the various interfaces between data, infrastructure and applications matters because the technical shape of data is formed in relation to the platform, and is indeed situated within a production context (e.g., Vis, 2013). Through an investigation of commensuration, it is clear that not all signals are equal, even if they can be counted, recombined or decontextualised. Commonness and similarity are not properties inherent to a metric, but rather constitute an accomplishment, indeed the outcome of commensuration. The analytical and transactional operations through which such data points are made commensurable and countable at the same time also facilitate practical aims such as the efficient management of activities and practices including advertising, customer relationship management (CRM), or search result ranking. Finally, I propose to engage more deeply with economic theory to properly study online social networks as data assemblages. The point is not to study economic problems or activity *per se*, but rather to approach problems and phenomena deemed ‘social’ through these perspectives to understand how the ‘social’ is literally accounted for in different cases. Such an approach is attentive to the calculative practices that make the ‘social’ visible and measurable in the capacity of markets alongside multiple value axes, and takes calculative practices, accounting techniques, commensuration and other techniques as performative agencies enrolled in ‘technologies of government’ (Rose and Miller, 1992: 183), or ‘the mechanisms through which programs of government are articulated and made operable’ (Espeland and Stevens, 1998: 379). Taken together, the specificity of the relationships between data, methodology and techniques constitutes an object of study warranting further investigation.

Acknowledgements

I would like to thank Anne Helmond, Bernhard Rieder, Jan Teurlings, and the anonymous reviewers for their constructive critical comments on previous versions of the article manuscript.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. These and other concerns are reflected in a recommendation report by the Executive Office of President Barack Obama titled 'Big Data: Seizing Opportunities, Preserving Values' (published in May 2014), which is the outcome of a 90-day review of Big Data at the start of 2014.
2. Following its etymology as documented in the *Oxford English Dictionary*, 'accounting, *n.*', at least since the late 14th century, pertains to 'The action or process of reckoning, counting, or computing; numeration, computation, calculation; (also) an instance of this' (1a; since 1486), 'With *for*. The giving of a satisfactory explanation of; answering for' (2a; since 1720), or 'An explanation; *esp.* one given as a justification for conduct, etc.' (2b; since 1885). These definitions enrich our understanding of the notion significantly, since the term is now chiefly used interchangeably with *accountancy* in a narrow sense, 'The action, process, or art of keeping and verifying financial accounts' (2b; since 1703).
3. Although in conventional accounting economic entities are usually considered to be institutions (e.g., hospitals, companies, agencies, states), a more general description may include any organisation or unit of society. As such, economic entities are those engaging in identifiable economic activities, controlling (or involved in the control of) economic resources, and which distinguish themselves from other entities and can be presented together (e.g., compared or consolidated).
4. There have been several historical and empirical studies of accounting in the specific contexts in which it operates (Burchell et al., 1985; Burns and Scapens, 2000; Hopwood, 1972, 1983, 1987; Hopwood and Miller, 1994; Miller, 1998).
5. Available at: <http://github.com/facebookarchive/scribe> (accessed 8 February 2015).
6. By means of illustrating the 'double social life' (Law et al., 2011) of this feature update, this new *index* is not just instrumental to retrieving posts, but also comes to constitute a new notion of group memory based on 'weaving different lives together by virtue of the words chosen to describe them' (Constine, 2014c), and greatly expands analytics possibilities based on post content (e.g., for generating suggestions of events, friends, pages, or target ads based on keywords). The implications of this update are therefore enormous.
7. Available at: <http://research.facebook.com/ai> (accessed 8 February 2015).
8. *Deep analytics* is loosely defined in industry as the application of sophisticated data processing techniques to extract information from large and typically multi-source data sets comprised of both 'unstructured' and 'semi-structured' data (as in Big Data), and which is often coupled with or part of query-based search mechanisms used in business intelligence or data mining applications (Janssen, n.d.; Rouse, 2011). Unlike 'structured' data, which is typically managed using a structured query language like SQL, unstructured data is those that cannot be so readily classified (e.g., images, video, audio, PDF files, emails, webpages, blog entries or wiki pages). Semi-structured data is in-between these two. It is a type of structured data, but lacks the strict data model structure.
9. Using a variety of techniques such as natural language processing, text analysis, computational linguistics, text mining or dealing with unstructured data, new practices have emerged that utilise social media services to do *social analytics*. Examples of this include sentiment analysis and opinion mining, customer experience management, social Customer Relationship Management (CRM), market research techniques like 'voice of the customer' to identify customer sentiment, trends or needs or using Klout scores in decision-making (e.g., Gerlitz and Lury, 2014).
10. Facebook – following competitors like Google and Microsoft – has taken up and begun investing heavily in *deep learning*, an emerging AI technique that is faster and more efficient than conventional forms of machine learning. It achieves this by simulating neural networks to process data and thereby pushes the limit of what can be implemented with this technique; for example, 'helping people organise their photos or choose which is the best one to share on Facebook', or 'prune the 1,500 updates that average Facebook users could possibly see down to 30 to 60 that are judged most likely to be important to them' (Simonite, 2013).
11. This involves such things as the distribution of venture capital funds or hiring practices (e.g., buying companies for their engineers) by Facebook, Google, and others. Facebook, for example, hired deep learning expert Marc'Aurelio Ranzato away from Google for its new AI Research lab. This group also includes: Yaniv Taigman, who is a cofounder of Face.com; computer vision expert Lubomir Bourdev; and Keith Adams, a veteran Facebook engineer (Simonite, 2013).
12. Moritz Hardt provides race and gender as examples of attributes that 'are typically redundantly encoded in any sufficiently rich feature space whether they are explicitly present or not' (2014). They are latent in the observed attributes, and this is why learning algorithms are able to pick up on these statistical patterns in training data. The far-reaching implications of such pervasive profiling models have led Richard Mortier et al. to write a manifesto-like article concerned with 'human–data interaction', arguing for human–data interaction that is characterised by 'negotiability', and building ethical systems for a data-driven society (Haddadi et al., 2013; Mortier et al., 2015).
13. This may include such things as considering the performative capacities of links, clicks, shares and social buttons in the creation of web economies (Gerlitz and Helmond, 2013) or the role of devices like search engines and social media platforms in the making of 'real-time' (Weltevrede et al., 2014).

References

- Agneeswaran VS (2014) *Big Data Analytics Beyond Hadoop: Real-Time Applications with Storm, Spark, and More Hadoop Alternatives*. Upper Saddle River, NJ: Pearson Education, Inc.
- Ahuja SP and Moore B (2013) State of Big Data analysis in the cloud. *Network and Communication Technologies* 2(1): 62–68.
- Aiyer A, Bautin M, Chen GJ, et al. (2012) Storage infrastructure behind Facebook messages using HBase at scale. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* 35(2): 4–13.
- Anderson C (2008) The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*, 23 June, 16(7). Available at: http://www.wired.com/science/discoveries/magazine/16-07/pb_theory (accessed 8 February 2015).
- Andrejevic M (2014) The Big Data divide. *International Journal of Communication* 8(10): 1673–1689.
- Apache Software Foundation (2013). *Hadoop 1.2.1 Documentation*. Available at: <http://hadoop.apache.org/docs/r1.2.1/> (accessed 8 February 2015).
- Argote L (2013) *Organizational Learning: Creating, Retaining and Transferring Knowledge*. Boston, MA: Springer US.
- Barocas S and Selbst AD (2014) Big Data's disparate impact. SSRN Working Paper Series, Social Science Research Network, USA, August. Available at: <http://ssrn.com/abstract=2477899> (accessed 8 February 2015).
- Becker BC and Ortiz EG (2009) Evaluation of face recognition techniques for application to Facebook. In: *FG '08 proceedings of the 8th IEEE international conference on automatic face & gesture recognition*, Amsterdam, the Netherlands, 17–19 September 2008, pp. 1–6. IEEE Xplore. DOI: 10.1109/afgr.2008.4813471.
- Beyer M (2011) Gartner says solving 'Big Data' challenge involves more than just managing volumes of data. *Gartner*. Available at: <http://www.gartner.com/newsroom/id/1731916> (accessed 2 September 2015).
- Bijker WE and Law J (eds) (1992) *Shaping Technology/ Building Society: Studies in Sociotechnical Change*. Cambridge, MA: The MIT Press.
- Bijker WE, Hughes TP and Pinch TJ (eds) (1987) *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA: The MIT Press.
- Borthakur D, Rash S, Schmidt R, et al. (2011) Apache Hadoop goes realtime at Facebook. In: *SIGMOD '11 proceedings of the 2011 ACM SIGMOD international conference on management of data*, Athens, Greece, 12–16 June 2011, pp. 1071–1080. New York: ACM Press.
- Borthakur D (2013) HDFS Architecture guide. In: Hadoop 1.2.1 documentation. Available at: http://hadoop.apache.org/docs/r1.2.1/hdfs_design.html (accessed 8 February 2015).
- boyd dm and Crawford K (2011) Six provocations for Big Data. In: *A decade in internet time: Symposium on the dynamics of the internet society*, Oxford, UK, 21–24 September 2011, pp. 1–17.
- boyd dm and Crawford K (2012) Critical questions for Big Data. *Information, Communication & Society* 15(5): 662–679.
- Bozkır AS, Güzin Mazman S and Akçapınar Sezer E (2010) Identification of user patterns in social networks by data mining techniques: Facebook case. In: Kurbanoglu S, Al U, Lepon Erdoğan P, et al. (eds) *Technological Convergence and Social Networks in Information Management, Communications in Computer and Information Science*. Vol. 96, Berlin: Springer-Verlag, pp. 145–153.
- Burchell S, Clubb C and Hopwood AG (1985) Accounting in its social context: Towards a history of value added in the United Kingdom. *Accounting, Organizations and Society* 10(4): 381–413.
- Burns J and Scapens RW (2000) Conceptualizing management accounting change: An institutional framework. *Management Accounting Research* 11(1): 3–25.
- Burrows R and Savage M (2014) After the crisis? Big Data and the methodological challenges of empirical sociology. *Big Data & Society* 1(1): 1–6.
- Callon M (1986) Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. In: Law J (ed.) *Power, Action and Belief: A New Sociology of Knowledge?* London: Routledge, pp. 196–223.
- Callon M (ed.) (1998) *The Laws of the Markets*. Oxford: Blackwell Publishers.
- Callon M and Law J (2005) On qualculation, agency, and otherness. *Environment and Planning D: Society and Space* 23(5): 717–733.
- Callon M and Muniesa F (2005) Peripheral vision: Economic markets as calculative collective devices. *Organization Studies* 26(8): 1229–1250.
- Callon M, Millo Y and Muniesa F (2007) *Market Devices*. Oxford: Wiley-Blackwell.
- Chayka K (2014) This is what your face looks like to Facebook: Artist Sterling Crispin's 'Data Masks' remind us the machines are always watching. In: Matter – medium. Available at: <https://medium.com/matter/this-is-what-your-face-looks-like-to-facebook-f771b3e11ed> (accessed 8 February 2015).
- Chu C-T, Kim SK, Lin Y-A, et al. (2006) Map-reduce for machine learning on multicore. In: *Advances in Neural Information Processing Systems*. Vol 19: pp. 306–313.
- Constine J (2014a) Facebook brings graph search to mobile and lets you find feed posts by keyword. In: TechCrunch. Available at: <http://techcrunch.com/2014/12/08/facebook-keyword-search/> (accessed 8 February 2015).
- Constine J (2014b) Hands-on with Facebook post search: Strong recommendations, yelp should worry. In: TechCrunch. Available at: <http://techcrunch.com/2014/12/08/every-post-is-an-opinion/> (accessed 8 February 2015).
- Constine J (2014c) The enormous implications of Facebook indexing 1 trillion of our posts. In: TechCrunch. Available at: <http://techcrunch.com/2014/12/28/mining-the-hive-mind/> (accessed 8 February 2015).

- Cortes C and Vapnik V (1995) Support-vector networks. *Machine Learning* 20(3): 273–297.
- Crawford K (2013) Think again: Big Data. In: Foreign Policy. Available at: http://www.foreignpolicy.com/articles/2013/05/09/think_again_big_data (accessed 8 February 2015).
- Crawford K (2014) The anxieties of Big Data. In: The New Inquiry. Available at: <http://thenewinquiry.com/essays/the-anxieties-of-big-data/> (accessed 8 February 2015).
- Crawford K and Schultz J (2014) Big Data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review* 55(1): 93–128.
- Day S, Lury C and Wakeford N (2014) Number ecologies: numbers and numbering practices. *Distinktion: Scandinavian Journal of Social Theory* 15(2): 123–154.
- Dean J and Ghemawat S (2008) MapReduce: Simplified data processing on large clusters. *Communications of the ACM* 51(1): 107–113.
- Dean J and Ghemawat S (2010) MapReduce: A flexible data processing tool. *Communications of the ACM* 53(1): 72–77.
- Derrida J (1982) *Positions*. Chicago: University of Chicago Press.
- Einav L and Levin J (2014) Economics in the age of big data. *Science* 346(6210): 1243089.
- Ekbia H, Mattioli M, Kouper I, et al. (2014) Big Data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology* 66(8): 1523–1545.
- Espeland WN and Sauder M (2007) Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology* 113(1): 1–40.
- Espeland WN and Stevens ML (1998) Commensuration as a social process. *Annual Review of Sociology* 24(1): 313–343.
- Etherington D (2014) Facebook's deepface project nears human accuracy in identifying faces. In: TechCrunch. Available at: <http://techcrunch.com/2014/03/18/faceook-deepface-facial-recognition/> (accessed 8 February 2015).
- Evans JR and Lindner CH (2012) Business analytics: The next frontier for decision sciences. *Decision Line* 43(2): 4–6.
- Feldman MS and March JG (1981) Information in organizations as signal and symbol. *Administrative Science Quarterly* 26(2): 171–186.
- Fligstein N (1993) *The Transformation of Corporate Control*. Cambridge, MA: Harvard University Press.
- Foucault M (1995) *Discipline and Punish: The Birth of the Prison*, trans. Sheridan AM. New York, NY: Vintage.
- Gerlitz C and Helmond A (2013) The Like economy: Social buttons and the data-intensive web. *New Media & Society* 15(8): 1348–1365.
- Gerlitz C and Lury C (2014) Social media and self-evaluating assemblages: On numbers, orderings and values. *Distinktion: Scandinavian Journal of Social Theory* 15(2): 1–15.
- Ghemawat S, Gobiuff H and Leung S-T (2003) The Google file system. In: *SOSP '03 proceedings of the 19th ACM symposium on operating systems principles*, Bolton Landing, NY, USA, 19–22 October 2003, pp. 29–43. New York: ACM.
- Gitelman L (ed.) (2013) *'Raw Data' Is an Oxymoron*. Cambridge, MA: The MIT Press.
- Goody J (1986) *The Logic of Writing and the Organization of Society*. Cambridge: Cambridge University Press.
- Govindaraju V, Rajan SP and Bhardwaj A (n.d.) Meta machine learning deep analytics on 'Bigger Data'. *Center for Unified Biometrics and Sensors*. Available at: http://www.cubs.buffalo.edu/Meta_Machine_Learning_Deep_Analytics_on_Bigger_Data.html (accessed 8 February 2015).
- Granovetter M (1985) Economic action and social structure: The problem of embeddedness. *American Journal of Sociology* 91(3): 481–510.
- Haddadi H, Mortier R, McAuley D, et al. (2013) Human-data interaction. Technical Report No. 837. Cambridge: University of Cambridge Computer Laboratory.
- Hardt M (2014) How big data is unfair: Understanding sources of unfairness in data driven decision making. In: Medium. Available at: <https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de> (accessed 8 February 2015).
- Hicks M (2011) See the messages that matter. In: Facebook. Available at: <https://www.facebook.com/notes/facebook/see-the-messages-that-matter/452288242130> (accessed 8 February 2015).
- Hopwood AG (1972) An empirical study of the role of accounting data in performance evaluation. *Journal of Accounting Research* 10(Suppl): 156–182.
- Hopwood AG (1983) On trying to study accounting in the context in which it operates. *Accounting, Organizations and Society* 28(2–3): 287–305.
- Hopwood AG (1987) The archeology of accounting systems. *Accounting, Organizations and Society* 12(3): 207–234.
- Hopwood AG and Miller P (1994) *Accounting as Social and Institutional Practice*. Cambridge: Cambridge University Press.
- Hu H, Wen Y, Chua T-S, et al. (2014) Toward scalable systems for Big Data Analytics: A technology tutorial. *IEEE Access* 2: 652–687.
- Janssen C (n.d.) Deep analytics. In: Techopedia. Available at: <http://www.techopedia.com/definition/26593/deep-analytics> (accessed 8 February 2015).
- Keat R and Urry J (2011) *Social Theory as Science*. London: Taylor & Francis.
- Ketelaar E (2006) 'Control through Communication' in a comparative perspective. *Archivaria* 60: 71–89.
- Kitchin R (2014a) Big Data, new epistemologies and paradigm shifts. *Big Data & Society* 1(1): 1–12.
- Kitchin R (2014b) *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. London: SAGE Publications.
- Kohavi R and Provost F (1998) Glossary of terms. *Machine Learning* 30(2–3): 271–274.
- Krishnan K (2013) *Data Warehousing in the Age of Big Data*. Amsterdam: Elsevier Science.
- Latour B (1987) *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.
- Latour B and Woolgar S (1986) *Laboratory Life: The Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.

- Law J, Ruppert E and Savage M (2011) The double social life of methods. CRESC Working Paper Series No. 95, Centre for Research on Socio-Cultural Change, Open University, UK, March. Available at: <http://research.gold.ac.uk/7987/> (accessed 8 February 2015).
- Lindsay R (2009) Distributed data analysis at Facebook. In: Facebook. Available at: <https://www.facebook.com/notes/facebook-data-science/distributed-data-analysis-at-facebook/114588058858> (accessed 2 September 2015).
- Litterer JA (1959) *The Emergence of Systematic Management As Shown by the Literature of Management from 1870 to 1900*, PhD Thesis, University of Illinois, Champaign–Urbana, IL.
- Lustig I, Dietrich B, Johnson C, et al. (2010) The analytics journey. *Analytics*, November–December, pp. 11–18. Available at: <http://www.analytics-magazine.org/november-december-2010/54-the-analytics-journey> (accessed 8 February 2015).
- McAfee A and Brynjolfsson E (2012) Big Data: The management revolution. *Harvard Business Review* 90(10): 60–68.
- Macho T (2003) Zeit und Zahl: Kalender- und Zeitrechnung als Kulturtechniken. In: Krämer S and Bredekamp H (eds) *Bild – Schrift – Zahl*. Munich: Fink, pp. 179–192.
- Macho T (2008) Tiere zweiter Ordnung: Kulturtechniken der Identität und Identifikation. In: Baecker D, Kettner M and Rustemeyer D (eds) *Über Kultur Theorie und Praxis der Kulturreflexion*. Bielefeld: Transcript Verlag, pp. 99–117.
- Mackenzie A (2011) More parts than elements: How databases multiply. *Environment and Planning D: Society and Space* 30(2): 335–350.
- Mackenzie A (2014) Multiplying numbers differently: An epidemiology of contagious convolution. *Distinktion: Scandinavian Journal of Social Theory* 15(2): 189–207.
- Mackenzie A (2015) The production of prediction: What does machine learning want? *European Journal of Cultural Studies* 18(4–5): 429–445.
- Mackenzie A and McNally R (2013) Living multiples: How large-scale scientific data-mining pursues identity and differences. *Theory, Culture & Society* 30(4): 72–91.
- MacKenzie DA (2007) *Do Economists Make Markets? On the Performativity of Economics*. Princeton, NJ: Princeton University Press.
- MacKenzie DA and Wajzman J (1985) *The Social Shaping of Technology: How the Refrigerator Got Its Hum*. Milton Keynes: Open University Press.
- Mair M, Greiffenhagen C and Sharrock WW (2015) Statistical practice: Putting society on display. *Theory, Culture & Society* 0(0): 1–27 Epub ahead of print 6 January 2015. DOI: 10.1177/0263276414559058.
- Manovich L (2012) Trending: The promises and the challenges of big social data. In: Gold MK (ed.) *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, pp. 460–475.
- Mayer-Schönberger V and Cukier K (2013) *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston, MA: Houghton Mifflin Harcourt.
- Menon A (2012) Big Data @ Facebook. In: *MBDS '12 proceedings of the 2012 workshop on management of big data systems*, San Jose, CA, 21 September 2012, pp. 31–32. New York: ACM.
- Miller P (1998) The margins of accounting. *European Accounting Review* 7(4): 605–621.
- Miller P (2001) Governing by numbers: Why calculative practices matter. *Social Research* 68(2): 379–390.
- Minelli M, Chambers M and Dhiraj A (2012) *Big Data, Big Data Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*. Hoboken: John Wiley & Sons, Inc.
- Moretti F (2013) *Distant Reading*. London: Verso Books.
- Mortier R, Haddadi H, Henderson T, et al. (2015) Human–data interaction: The human face of the data-driven society. *Computing Research Repository*. Available at: <http://arxiv.org/abs/1412.6159> (accessed 8 February 2015).
- Porter TM (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.
- Power M (2004) Counting, control and calculation: Reflections on measuring and management. *Human Relations* 57(6): 765–783.
- Provost F and Kohavi R (1998) Guest editors' introduction: On applied research in machine learning. *Machine Learning* 30(2–3): 127–132.
- Richards NM and King JH (2013) Three paradoxes of Big Data. *Stanford Law Review Online* 66(41): 41–46.
- Richards NM and King JH (2014) Big Data ethics. *Wake Forest Law Review* 49: 393–432.
- Robson K (1992) Accounting numbers as 'inscription': Action at a distance and the development of accounting. *Accounting, Organizations and Society* 17(7): 685–708.
- Robson K (1994) Inflation accounting and action at a distance: The Sandilands episode. *Accounting, Organizations and Society* 19(1): 45–82.
- Rose N and Miller P (1992) Political power beyond the state: Problematics of government. *The British Journal of Sociology* 43(2): 173–205.
- Rouse M (2011) Deep analytics. March 2011. In: *SearchBusinessAnalytics – TechTarget*. Available at: <http://searchbusinessanalytics.techtarget.com/definition/deep-analytics> (accessed 8 February 2015).
- Ruppert E (2012) The governmental topologies of database devices. *Theory, Culture & Society* 29(4–5): 116–136.
- Ruppert E (2013) The economies and ecologies of Big Data. *New Social Media, New Social Science (NSMNSS) Digital Debate*, 16 April. Available at: <https://www.youtube.com/watch?v=hIFD118zJCo> (accessed 8 February 2015).
- Saitta L and Neri F (1998) Learning in the 'Real World'. *Machine Learning* 30(2–3): 133–163.
- Sauder M and Espeland WN (2009) The discipline of rankings: Tight coupling and organizational change. *American Sociological Review* 74(1): 63–82.
- Savage M (2010) *Identities and Social Change in Britain since 1940: The Politics of Method*. Oxford: Oxford University Press.
- Savage M (2013) The 'social life of methods': A critical introduction. *Theory, Culture & Society* 30(4): 3–21.
- Savage M and Burrows R (2007) The coming crisis of empirical sociology. *Sociology* 41(5): 885–899.

- Schmidt L-H (1996) Commonness across cultures. In: Balslev AN (ed.) *Cross-cultural Conversation (Initiation)*. Atlanta, GA: Scholars Press.
- Shvachko KV (2010) HDFS scalability: The limits to growth. *Login* 35(2): 6–16.
- Siebert B (2013) Cultural techniques: Or the end of the intellectual postwar era in German media theory. *Theory, Culture & Society* 30(6): 48–65.
- Siebert B (2014) *Cultural Techniques: Grids, Filters, Doors, and Other Articulations of the Real*. New York, NY: Fordham University Press.
- Simonite T (2014) Facebook creates software that matches faces almost as well as you do. In: MIT technology review. Available at: <http://www.technologyreview.com/news/525586/facebook-creates-software-that-matches-faces-almost-as-well-as-you-do/> (accessed 8 February 2015).
- Simonite T (2013) Facebook launches advanced AI effort to find meaning in your posts.. In: MIT technology review. Available at: <http://www.technologyreview.com/news/519411/facebook-launches-advanced-ai-effort-to-find-meaning-in-your-posts/> (accessed 8 February 2015).
- Star SL (ed.) (1995) *Ecologies of Knowledge: Work and Politics in Science and Technology*. Albany, NY: SUNY Press.
- Stone Z, Zickler T and Darrell T (2008) Autotagging Facebook: Social network context improves photo annotation. In: *CVPRW '08. IEEE computer society conference on computer vision and pattern recognition workshops*, Anchorage, UK, 23–28 June 2008, pp. 1–8. IEEE Xplore. DOI: 10.1109/cvprw.2008.4562956.
- Taylor L, Schroeder R and Meyer E (2014) Emerging practices and perspectives on Big Data analysis in economics: Bigger and better or more of the same? *Big Data & Society* 1(2): 1–10.
- Thusoo A, Sarma JS, Jain N, et al. (2009) Hive: A warehousing solution over a map-reduce framework. *Proceedings of the VLDB Endowment* 2(2): 1626–1629.
- Thusoo A, Sarma JS, Jain N, et al. (2011) Hive – A petabyte scale data warehouse using hadoop. In: *2010 IEEE 26th international conference on data engineering (ICDE)*, Long Beach, CA, 1–6 March 2010, pp. 996–1005. IEEE Xplore. DOI: 10.1109/icde.2010.5447738.
- Thusoo A, Shao Z, Anthony S, et al. (2010) Data warehousing and analytics infrastructure at Facebook. In: *SIGMOD '10 proceedings of the 2010 ACM SIGMOD international conference on management of data*, Indianapolis, IN, 6–11 June 2010, pp. 1013–1020. New York: ACM Press.
- Vis F (2013) A critical reflection on Big Data: Considering APIs, researchers and tools as data makers. *First Monday* 18(10) DOI: 10.5210/fm.v18i10.4878.
- Weltevrede EJT, Helmond A and Gerlitz C (2014) The politics of real-time: A device perspective on social media platforms and search engines. *Theory, Culture & Society* 31(6): 125–150.
- Woolgar S (1991) The turn to technology in social studies of science. *Science, Technology, & Human Values* 16(1): 20–50.
- Yang H-C, Dasdan A, Hsiao R-L, et al. (2007) Map-reduce-merge: Simplified relational data processing on large clusters. In: *SIGMOD '07 proceedings of the 2007 ACM SIGMOD international conference on management of data*, Beijing, China, 12–14 June 2007, pp. 1029–1040. New York: ACM Press.
- Yates J (1993) *Control through Communication: The Rise of System in American Management*. Baltimore: Johns Hopkins University Press.