Original Research Article

BIG DATA & SOCIETY

Big Data & Society
July–December: 1–13

DOI: 10.1177/20539517211040760
journals.sagepub.com/home/bds
SAGE

Discovering needs for digital capitalism: The hybrid profession of data science

Robert Dorschel

Abstract

Over the last decade, 'data scientists' have burst into society as a novel expert role. They hold increasing responsibility for generating and analysing digitally captured human experiences. The article considers their professionalization not as a functionally necessary development but as the outcome of classification practices and struggles. The rise of data scientists is examined across their discursive classification in the academic and economic fields in both the USA and Germany. Despite notable differences across these fields and nations, the article identifies two common subjectivation patterns. Firstly, data scientists are constructed as hybrids, who combine generally conflictive roles as both generalists and specialists; technicians and communicators; data exploiters and data ethicists. This finding is interpreted as demonstrating a discursive distinction between data scientists and other competing and supposedly more one-dimensional professionals, such as statisticians or computer scientists. Secondly, the article uncovers a discursive construction that interpellates data scientists as discoverers of needs. They are imagined as explorative work subjects who can establish growth for digital capitalism by generating behavioural patterns that allow for personalization, customization and optimization practices.

Keywords

Data scientist, tech worker, discourse analysis, professionalization, Bourdieu, Foucault

Introduction

Digital technologies and algorithms increasingly influence our decisions around the news we read, the products we buy, the people we trust, and even the people we love. Social scientists have provided insights into how processes of datafication and quantification are embedded, through elective affinity, in a digital capitalist system that relies heavily on proprietary platform ecosystems and surveillance practices (Couldry and Mejias, 2019; Fuchs, 2013; Mau, 2019; Sadowski, 2019; Staab and Nachtwey, 2016; Zuboff, 2019). Yet we still know surprisingly little about the professionals within the 'tech industry' (Tarnoff and Weigel, 2020) who generate and shape the technologies that permeate and influence social life (Wajcman, 2018). By contrast, the labour of precarious occupational groups, such as food delivery couriers or Uber drivers (Graham et al., 2017; Lehdonvirta, 2018), has been much more extensively analysed. This article suggests, however, that while technological advancement in capitalism may lead to deskilling (Braverman, 1998) or precarisation (Crouch, 2019), it also generates opportunities for new white-collar occupations (Shestakofsky, 2017).

This paper focuses on data scientists, an emerging group of tech professionals increasingly responsible for capturing, translating and commodifying human experiences through and into digital technologies, performing key operations for so-called 'surveillance capitalism' (Zuboff, 2019), 'data colonialism' (Couldry and Mejias, 2019) or 'program earth' (Gabrys, 2016). Data science originally emerged in the United States and has since been institutionalized within various contexts worldwide (Brandt, 2016; González-Bailón, 2017; Hammerbacher, 2009; Kotras, 2020; Metcalf and Crawford, 2016; Slota et al., 2020). Yet data is not new, and even increased quantities of digital data do not automatically call for the creation of a new occupation to handle large-scale data. The emergence of data science poses a puzzle since we can assume that in principle, existing professionals such as statisticians or computer scientists could also be qualified or trained to occupy the jurisdiction of expert digital data work. Instead of considering the professionalization of data

[1]Department of Sociology, University of Cambridge, UK

Corresponding author:
Robert Dorschel, Department of Sociology, University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, UK.
Email: rcd49@cam.ac.uk

science as a functionally necessary development, I will thus address its emergence as a process that relies on symbolic classification practices and struggles. In this article, I will analyse how the professional role of data scientists is discursively constructed by the academic and economic fields. I will not reconstruct the discursive construction of data science as a discipline. Rather, I will inquire into what kind of professional subjects find themselves increasingly in charge of the data work jurisdiction within digital capitalism. This article will explore how the profession of data science is imagined and understood externally, through an analysis of the construction of data scientists across the discourses of the academic and economic fields. These fields can be considered key social spaces for jurisdictional claims (Abbott, 1988: 55).

The few important studies thus far on data science as an occupation have mainly taken an inside perspective, ethnographically studying or interviewing data scientists themselves in order to reconstruct the story of the rise of the profession (Knox and Nafus, 2018; Mützel et al., 2018; Passi and Sengers, 2020; Slota et al., 2020). Abbott provides a framework for understanding data science as an emerging profession that has managed to gain control over a privileged task jurisdiction within a competitive 'system of professions' (Abbott, 1988). Drawing on Bourdieu, we may further understand this system of professions as embedded within a field of power where different social fields engage in classification struggles (Bourdieu, 1996b). Through this lens, we can consider professions as professional fields that engage in struggle via different forms of capital in a multidimensional space. However, both Abbot and Bourdieu rarely make use of the approach of discourse analysis. In their frameworks, discourse is merely a resource for professions (Abbott, 1988: 59; Bourdieu, 1991; Bourdieu and Wacquant, 1992: 105). A discursive theoretical approach regards professions as effects of discourse, and aims to analyse the knowledge and power relationships inherent in and sustained through discourse (Foucault, 2002). It is especially important to explore professions in the making through this lens, I argue, because discourses allow for breaks in society. Discourses enfold subjectivation effects, meaning they modulate the dynamic knowledge orders according to which subjects and collectives are classified and self-classify (Foucault, 2005).

This article analyses the discursive construction of data scientists across the economic and academic fields in the United States and Germany. I focus on academic study programs and job ads. While the discursive constructions within these text materials cannot be taken at face value as representing the logics of non-discursive practice, they do possess a high degree of visibility and can be considered obligatory passing points for data scientists (Ingram and Allen, 2019). The United States is the country where the profession of data scientist first appeared and, to date, it represents the largest digital economy in the world, while Germany is the largest digital economy in Europe (UNCTAD, 2019). The study finds striking commonalities between the construction and classification of data scientists within these two leading digital economies. However, there are also notable differences present, which I will use to provide a sense of national varieties of data (science) imaginations. Methodologically, I use the sociology of knowledge approach to discourse analysis (Keller, 2011), which, based on Foucauldian discourse theory, provides a set of heuristics and procedures that can be used for effective qualitative analysis of the phenomenal structures, interpretation patterns and subjectivation forms within discourse.

My central finding is that the discursive construction of data scientists should be understood primarily by means of their classification as *hybrid work subjects,* capable of *discovering new needs*. In both academic and economic fields, big data is interpreted as a resource which data scientists can mine for new insights into customer needs as well as commercial production processes. Data scientists are discursively understood to have the capacity to detect and even predict needs through their analysis of behavioural patterns and programming of algorithms. In this way, data scientists are imagined as work subjects who can establish new growth for the capitalist system through a combination of personalized and customized digital products and services and the optimization of production processes at the firm level.[1] Authors such as Mackenzie (2018) have argued that the promise of personalization underpins big data quantification. Extending their insight, I demonstrate how this promise primarily operates through the semantic of needs.

The second interconnected finding is the discursive construction of data scientists as hybrids. Despite notable differences between the United States and Germany, in both territories, data scientists are discursively constructed as professional subjects who combine several conflictive skills sets and character traits. The academic and economic fields interpellate data scientists as work subjects with habitual dispositions that are traditionally assigned to distinct social figures. Data scientists are simultaneously constructed as *generalists and specialists; technicians and communicators; data exploiters and data ethicists*. I interpret this form of hybrid subjectivation as a strategic advantage granted to data scientists vis-à-vis supposedly more one-dimensional occupations and professions such as statisticians or computer scientists (Kendall, 1999), who also compete over the jurisdiction of digital data work. This assumed hybridity allows data scientists to integrate the different expectations deemed necessary to fulfil the task of digital need-discovery.[2] In theoretical terms, this finding does not point to hybridity as a general habitus of modern subjects (e.g. Lahire, 2011), but rather as an unequally distributed form of cultural capital that is produced in discourse. Thus, while previous research argues that the rise of data scientists is based on their capacity to render data

from worldly domains available for their operations (Slota et al., 2020), I argue that the process rests at an upstream level on the strategic subjectivation of data scientists as hybrids who can discover new needs for digital capitalism.

The discursive making of a professional field

For Abbott, professionals do not necessarily fulfil more important or more complex societal tasks than non-professional, less autonomous workers. In opposition to functional scholars of the professions (e.g. Parsons, 1939), he argues that professional status emerges primarily through power struggles and social acts of construction. Abbott uses the concept of jurisdiction, within his ecological framework, to analyse professionalization strategies (Abbott, 1988: 59, 2005: 248). Here, professions are considered to be in constant competition with each other over jurisdictional task areas, and aim to secure or expand their territory through the construction and presentation of abstract knowledge expertise and skills (Abbott, 1988: 55). Such jurisdictional disputes are always relational processes: a jurisdictional expansion for some is a defeat for others (Abbott, 1988: 91, 143).

Abbott is primarily concerned with conflicts between existing professions, but data science can be considered a new, emerging profession that is increasingly occupying the changing jurisdiction of expert data work. Data scientists can be thought of as 'corporate professionals' (Muzio et al., 2011), which means they secure their privileged positions within organizational contexts. Their emergence poses a puzzle, however, since we can assume that in principle, existing professionals such as statisticians or computer scientists might also be qualified or trained to occupy the jurisdiction of expert digital data work. To better understand the institutionalization of data science, I propose to turn to the embeddedness of professions within 'the field of power' (Bourdieu, 1996b). Drawing on Bourdieusian field theory, we can understand professions as 'professional fields' whose positioning in social space is tied to the actions and strategies of other social fields – a dimension that Abbott's seminal work fails to address.[3] Fields refer to differentiated spheres with a relatively autonomous symbolic order that expresses itself in specific rules and logics (Bourdieu, 1996a). Within social fields, actors compete over positions of authority by using different forms of capital. This struggle is oriented towards classifying legitimate rules and thereby the legitimate habitus, understood as a network of schemas of perception, recognition and action (Bourdieu, 1990: 53). On the other hand, social fields are themselves embedded in a field of power, a meta-field, where fields struggle over the nomos of social space by imposing classification systems on other fields (Bourdieu, 1996b: 336). Through this Russian-doll-like heuristic, data science can be considered a proto-field which is embedded in (without being determined by) the relations and struggles within the field of power.

However, both Abbott's and Bourdieu's theories consider discourse primarily as a resource of power rather than as a self-dynamic social structure. Through their lens, professions or professional fields are figures of domination that hold discursive power (Abbott, 1988: 59; Bourdieu, 1991; Bourdieu and Wacquant, 1992: 105). In contrast, Foucault holds that discourses are neither determined by interests nor merely representative of phenomena (Foucault, 2002). While Foucault's theory is also primarily concerned with power relations in social life, he considers discourse a self-dynamic force, conceptualized as a system of utterances that exerts subjectivation effects and that performatively constructs the subjects and worlds of which it speaks. At stake in a discourse is the fixing of collective knowledge orders that determine the grounds on which subjects are classified and self-classify (Keller, 2011: 48).[4] Discourses are made possible and ordered on the basis of relatively autonomous formation rules, whose logic cannot be derived deterministically from the speakers or their interests (Foucault, 2002: 42). The goal of discourse analysis is to unearth the formation rules as well as the power/knowledge nexus that underlies discursive complexes like the discourse around data science.[5] Through the lens of discourse theory, then, professions are not in the first instance figures of domination, but rather they act as links within a set of power-knowledge relations. Furthermore, the elective affinity between the profession of data science and digital capitalism must also be understood as a result of relatively self-dynamic discursive relations. The Foucauldian lens is especially useful in guiding us towards breaks and changes in society, such as the making of a profession. (Goldstein, 1984). For instance, Foucault demonstrates how discourses around modern medicine not only changed social understandings of bodies but also brought about new 'disciplines', as well as disciplined and disciplinary-trained professional subjects, including modern medical practitioners with their medical gaze (Foucault, 2003: 80, 89).

Recent developments around digital technologies confront us with dynamic ensembles of human and non-human actors that have produced new task areas (Blok et al., 2019; Gabrys, 2011: 48). Data scientists constitute one emergent professional group that has been able to benefit from datafication developments and may be expected to continue increasing in rank, given the expansion of the digital economy in the wake of the COVID-19 pandemic. So far, data scientists have received a fair amount of journalistic attention (e.g. Davenport and Patil, 2012), and have themselves tried to explain (and further strengthen) their emergence (Hammerbacher, 2009; O'Neil and Schutt, 2013). Yet these accounts often offer explanations centred on technical skills. The number of social scientific studies on data

science is moderately increasing but remains rather low (Brandt, 2016; González-Bailón, 2017; Knox and Nafus, 2018; Kotras, 2020; Passi and Sengers, 2020; Saner, 2019; Slota et al., 2020). The few existing studies tend to take an inside perspective, following the practices of data scientists themselves in their work and quest for institutionalization. I will offer an alternative inquiry into the professionalization of data scientists through an analysis of their discursive construction in the economic and academic fields. Instead of continuing the important endeavour of making sense of data scientists and their practices through ethnographic study or through interviewing, I will focus on their institutional interpellation. I will inquire into the subjectivation of data scientists through an analysis of their external discursive construction in the academic and economic fields. For Bourdieu, the economic field consists of the ensemble of actors who engage in trades of material and symbolic goods (Bourdieu, 2005), while the academic field consists of the ensemble of actors who engage in the production of institutional cultural goods (Bourdieu, 1988). Given the dominant positions of these two fields within the 'field of power' (Bourdieu and Wacquant, 1992: 229), understood as the space where fields struggle over defining the hierarchical relations between them, these two spheres can be considered key consecration instances for processes of professionalization.

Methodology and data

Discourse theory is realized through the corresponding methodology of discourse analysis, but Foucault never developed a detailed guideline of how his discourse theoretical research program should be operationalized. Thus, I draw on the sociology of knowledge approach to discourse analysis for my empirical research (SKAD) (Keller, 2011). The sociology of knowledge tradition builds explicitly on Foucault's interest in knowledge as the entirety of symbols, signs and cultural schemas, beyond what is typically socially recognized as knowledge. Like Bourdieu, this tradition is primarily interested in how discourse unfolds within the context of differentiated social spheres. SKAD offers two central conceptual heuristics for the analysis of discourses: firstly, phenomenal structure, which refers to the content-related structuring of discourse, and secondly, interpretation patterns, which depict fundamental meaning- and action-generating schemas that circulate through discourses and enfold subjectivation effects (Keller, 2011: 57–59).

Table 1. Data science master study programs.

USA	Germany
High-ranked universities:	*High-ranked universities:*
Brown University	Freie University Berlin
Columbia University	LMU Munich
Duke University	TU Munich
Harvard University	RWTH Aachen
Stanford University	University of Mannheim
Medium-ranked universities:	*Medium-ranked universities:*
Boston University	University of Bielefeld
Carleton University (Canada)	University of Jena
Georgetown University	University of Potsdam
Northeastern University	University of Leipzig
University of Virginia	University of Marburg
Low-ranked universities:	*Low-ranked universities:*
DePaul University	Leuphana University
Illinois Tech Institute	TU Braunschweig
Oklahoma University	TU Chemnitz
San Francisco University	TU Dortmund
Southern Methodist University	University of Saarland

Broadly following this analytical procedure, all textual materials examined in this article were first read and coded with regard to content-related themes in order to grasp their phenomenal structure. Memos were used to capture initial theoretical thoughts, narrative elements and general impressions. In a second step, the texts were re-read several times and more interpretative codes were applied. Building on this detailed analysis, codes were finally subsumed under general categories and abstracted to interpretation and subjectivation patterns. The textual data ($n = 110$ documents) is composed of data science study programs and job ads published between March 2019 and May 2020. While the discursivations within study programs and job ads cannot be taken at face value as representing the logics of non-discursive practice, both mediums can be considered 'obligatory passing points' (Callon, 1984: 216) for the development of data scientists. These materials are also highly visible and generate insights into how institutions imagine ideal graduates and elite professionals in the field (Ingram and Allen, 2019). While the United States is the country with the largest digital economy and is home to the first occupational popularization of data science, Germany is the largest European digital economy and at the forefront of second movers in the field. I will at times highlight significant differences in the discursive constructions between the United States and Germany, which can be interpreted against the backdrop of their embodiment of different market economies (Hall and Soskice, 2001). The USA represents an ideal-typical liberal market economy, while Germany stands for a coordinated market economy with a stronger industrial sector. The main focus of this paper, however, will lie on the striking commonalities between the patterns of discursivation in both national fields. Analysing the similar discursive construction of data scientists in these two different market economies allows for a tentative diagnosis of the typical features attending the institutionalization of data scientists in Western post-industrialized economies.

The empirical data was gathered with respect to the differentiated power structure of academic and economic

fields. For the academic field, whose most important actors are the universities according to Bourdieu (1988: 128), I collected materials related to study programs ($n = 30$) from five high-ranked universities, five medium-ranked universities and five low-ranked universities in both the United States and Germany (see Table 1).[6] I selected universities that included detailed descriptions of the aims and goals of the study programs as well as information on their curricula. For the economic field, whose central actors are firms according to Bourdieu (2005: 75), I collected job ads ($n = 80$) from the popular online platform glassdoor. I selected job ads that included a detailed description of both the work responsibilities and the necessary qualifications expected of applicants.[7] I sampled 40 job ads from established tech firms and startup tech firms in Berlin and the Bay Area (see Table 2). Only tech firms were used (and not, for instance, companies in other industries, such as traditional automobile firms) so as to focus on one central sector in demand of data science expertise. This focus presents a limitation of this study since the professional field of data science has institutionalized far beyond the field of tech. Governmental agencies, for instance, have also become a habitat for data scientists. Furthermore, an encompassing analysis of data scientists' discursive construction in the academic and economic field would require analysis of more types of documents, such as blogs, podcasts or firm-intern training modules. Last but not least, another limitation of this study is that it does not analyse the discursive construction of competing professions, such as statisticians or computer scientists.

Table 2. Data science jobs ads (number of analysed materials in parentheses).

USA	Germany
Established tech firms:	*Established tech firms:*
Amazon (4)	EVENTIM (4)
Apple (4)	Zalando (4)
Google (4)	WOOGA (2)
Microsoft (4)	Scout24 (2)
	Delivery Hero (2)
	HelloFresh (2)
Startup tech firms:	Startup tech firms:
JUUL Labs (4)	TIER (3)
WeWork (4)	Deevio (2)
InstaCart (3)	ShareNow (2)
Stripe (3)	Eyeo (2)
DoorDash (3)	Zero to One Search (2)
Slack (3)	Remerge (1)
Discord (2)	Smart Steel (1)
Metromile (1)	Fliit Holding (1)
Blue Owl (1)	Omio (1)
	Regis 24 (1)
	Körber (1)
	Hundred 5.0 (1)
	Marley Spoon (1)
	Zattoo (1)
	Ultimate AI (1)
	Frontier Car Group (1)
	ORAYLIS (1)
	Searchtalent (1)

The hybrid professionalization of data scientists as discoverers of needs

I will now analyse the discourse on institutional expectations of data scientists as professionals. The central finding of the discourse analysis is that data scientists are constructed as hybrid work subjects, understood to have diverse skills and roles that allow them to discover new needs by making use of large-scale data. The subjectivation of the academic and economic field attributes conflictive roles to data scientists, interpellating habitual dispositions that traditionally belong to distinct social figures. In the following section, I will reconstruct the three interpellated forms of hybridity that relate to data scientists and demonstrate their implicit and explicit entanglement with the dimension of 'needs'.

Data scientists as generalists and specialists

By tracking data science discourse in both academic and economic fields, it becomes clear that the virtues of both generalists and specialists are imagined to be necessary for the practice of recognizing needs. Discursive practices across both fields oscillate between these two conflicting modes of subjectivation; indeed, they often co-exist within the same document. The *generalist* subjectivation is based on positioning data scientists as subjects with broad expertise that draws on an array of disciplines. Study programs in both the United States and Germany highlight the ongoing interdisciplinary training data scientists will receive. They promise an education that prepares students for professional data work through a pluralistic skill set:

> Data Science is the science of extracting knowledge and information from data and requires competencies in both statistical and computer-based data analysis. The elite program Data Science is an interdisciplinary program. (LMU Munich, 2020)

While this extract does not reveal specific dimensional needs, it does indicate the two most important disciplines for data scientists: computer science and statistics. Computational skills, especially programming expertise, are understood to prepare data scientists for basic work tasks such as data cleaning, data analysis techniques and data visualization. Education in statistics, on the other hand, is considered necessary to provide the mathematical knowledge required to make sense of the data. This

construction of data science expertise as grounded in statistics and computer science knowledge is echoed in the economic field, as evident in a job listing for the Berlin-based game development company Wooga:

> The Data Science team explores new methodologies and delves deep into every aspect of our data to identify opportunities and make a real impact on our business. (…) You will work on analytics projects involving multiple team members to understand, segment, predict and influence the behaviour of our players across our games portfolio (…)
>
> **About you**:
>
> - You have a degree in a highly quantitative field (e.g. statistics, economics, research, computer science).
> - You have professional working experience in data science and business analytics.
> - You are proficient in using Python for data analysis, and you can use SQL effectively to query large volumes of data. (WOOGA 2020, GER)

This ad presents data scientists as explorative work subjects who can not only 'understand' but also 'predict' and even 'influence' customer behaviour. I interpret these attributed capacities around behavioural patterns as belonging to the general categorization of data scientists as need-discoverers. The qualifications required for this task, statistical knowledge and computer skills, correspond to the multidisciplinary curricula of the data science study programs. The subjectivation of data scientists as generalists is present at both startups and established tech firms (notably, though, there is more emphasis on generalism within the job ads of startups). Furthermore, the subjectivation pattern manifests in the organizational embeddedness of data science as an academic field. With the exception of a few cases in the United States, courses in data science are not typically organized through an autonomous institute or school. Instead, data science studies are usually embedded within Computer Science or Statistics departments. However, the institutional organization of data science as a subject without a home turf is not typically represented as a weakness but rather as a strength, implying it is a meta-discipline. This narrative is further supported by discursive homologies between the discipline of data science and the economic structures of the wider tech industry: '[o]ur faculty represent the fundamental multidisciplinary nature of the big data industry' (University of San Francisco 2019).

While the discursive construction of data scientists as generalists is the dominant modus of subjectivation, it is counteracted, at times, through a *specialist* framing. These traditionally distinct roles jointly form one aspect of the subjectivation of data scientists. The cultural code of specialism occurs in multiple ways. For instance, in study programs, students are urged to choose between different pathways: 'Specialization is possible in the profile areas of Data Science in the Social Sciences, Data Science in the Life Sciences, and Data Science Technologies' (Free University Berlin, 2020). Such specialization is deemed advisable in order for data scientists to have a 'precise understanding of the field from which their data originates' (Technical University Braunschweig, 2020). Thus, one thematic strand of discourse encourages a differentiation of skills within the occupational field of data science. Sometimes, the conflictive relationship between generalized and specialized knowledge manifests in a single utterance: 'Experience the full range of the data science ecosystem and graduate as an expert in at least one analytical approach or branch of technology' (Duke University, 2019).

The social figure of the specialist also appears regularly in job listings. Especially at established tech firms, job ads often specify which specialized project groups successful applicants would be assigned to, and may even further subdivide the kinds of labour required by each project. The cultural code of specialism also surfaces in frequent demands to work on practices that require nuance like 'personalization' and 'customization'. These practices act as central moral underpinnings of big data quantification (Mackenzie, 2018) and surface through the semantic of needs: Companies express an interest in data scientists as specialists who can 'tailor our product to suit different needs' (Discord, GER). The objective of data science work is often defined as the production of an exact match between the needs of customers and the available products:

> What You'll Do
>
> - Query understanding and intent identification to improve search and quantify consumer demand.
> - Improve our search, recommendation, personalization and ranking algorithms (…).
> - Build econometric models for next generation pricing algorithms
>
> (DoorDash 2019c, USA).

Constructions of data scientists as need-discoverers go hand in hand with their subjectivation as hybrid work subjects. Data scientists must have an interdisciplinary education that specifically combines broad statistical training and computational skills with specialized knowledge (about a domain or a method like econometric models, for instance) to be able to understand consumers' individual demands and influence their personal consumption, which can be interpreted as a strategy for new capitalist growth (Zuboff, 2019: 254). Discovering new areas for growth and 'disruption' seems to require dynamic work subjects who can master different forms of expertise. In the discourse analysed, data scientists are constructed as suitable work subjects for this task jurisdiction on the basis of the hybridity attributed to them.

Data scientists as technicians and communicators

Sociology has traditionally understood technical expertise and social-communicative competence as two different skill sets which tend to be incorporated by different occupational actors (Riesman et al., 2001: 128). In the institutional discourses that objectify data scientists, however, these traditionally conflictive social figurations co-exist, sometimes even symbiotically:

> We're looking for someone who has the technical skills to surface insights quickly, and the interpersonal skills to communicate those insights in a way that persuades action. (DoorDash 2019a, USA)

This ad invokes a subject who holds technical but also communication skills and is thus capable of transforming data into action. Along these lines, we also encounter the interpellation of the data scientist as a more than technical subject: 'You'll have deep technical skills and a real passion for helping Slack make data-informed decisions' (Slack 2019a, USA). Data science is imagined as a profession that can not only harness data but also deploy it for further use.

In line with the classification of data scientists as *technicians*, study programs often refer to the methodical skills they are said to acquire as 'tools': 'students learn to collect, organize, analyse and visualize big data through suitable tools and methods' (University of Mannheim 2020). Throughout the various materials, data scientists are classified as work subjects who can tame massive amounts of digital data and produce insights on the needs of customers through use of specific tools and method:

> You love to develop a feel for databases and to conduct creative feature engineering and to try out different classes of algorithms. (…) Together with our team you will analyse concrete problems and the needs of our customers in order to develop perfectly fitting solutions. (Searchtalent, GER)

Again, data scientists are represented as work subjects who enable more personalized services and products. However, there is variation in how they are imagined as technicians: on the one hand, we can semantically detect an interpellation as industrial-engineering technicians who master methods like 'machine learning', while on the other hand, data scientists are imagined as technicians who resemble the craftsman type:

> Apple's Strategic Data Solutions (SDS) team is looking for a talented individual who is passionate about crafting, implementing, and operating analytical solutions that have direct and measurable impact to Apple and its customers. (Apple 2019d, USA)

In the United States discourse, the notion of craftmanship is evident at various points (see Sennett, 2008: 24 for a similar diagnosis with regard to Linux coders). In the German context, data scientists are more closely linked to industrial technicians, which should be considered in light of the greater role industrial production still plays in Germany. In either case, both cultural codings have as their vanishing point the social figure of the technician. They both point to a work subject who is problem-oriented and practical, and who can help to understand and influence customers by mastering a technical process.

The interpretation pattern of data scientists as technicians has a counterpart in the construction of data scientists as *communicators*. Within both the academic and economic spheres, the discourse is concerned with negating associations with technical 'nerdiness'. Especially in the study programs of elite universities, communication is presented as an integral part of the curricula, justified via the perception that '[c]learly communicating problems, ideas, data, analysis approaches, results, and recommendations for action are vital for career success in technology and science' (Georgetown 2019). Data scientists are curated as work subjects with social and empathic capabilities, capabilities that are the object of training through various measures including special seminars, group work sessions and team project competitions. Within the academic field, these training measures aim at subjectivating collaborative actors. Sometimes the discourse even invokes data scientists as novel types of leaders; 'shepherds' in Foucauldian terms. Marking a manifestation of pastoral power, they are interpellated as experts in subtly guiding the various self-governed subjects within the project teams of the digital economy. However, it is important to note that the teaching of communicative skills is more frequent in US academic institutions, which can be interpreted as constituting a competitive regional advantage with regard to human capital vis-à-vis Germany. Furthermore, the fact that higher-ranked universities place greater emphasis on teaching not only technical skills but also social skills can be interpreted as a strategy that allows for the reproduction of hierarchical relations between universities within the academic field.

In the job ads, most US and German firms explicitly demand social skills as well as technical expertise. This finding stands in contrast to popular images of tech workers, who are regularly considered as 'nerds' (see Kendall, 1999 or HBO's show 'Silicon Valley'). Moreover, the conflictive social knowledge relationship between technical engineering and communicative talent is resolved by imagining data scientists as two-dimensional work subjects:

> Your collaborative spirit and business acumen match your technical skills, as teamwork is the foundation of our success. (deevio 2020, GER)

Actors with technical expertise are no longer considered mutually exclusive from those with communicative and emotional abilities: data science is presented as the best of both worlds. Professional success seems to demand a combination of technical skills (working with things) and cultural learning (working with people). In other words, digital professionals must now strive to embody both Steves – engineer Steve Wozniak and business leader Steve Jobs. This hybrid subjectivation is again co-constructed with the professional proto-jurisdiction of digital need-discovery. The link between needs and the hybrid construction of data scientists is clear in the list of responsibilities published in a job ad for German e-scooter company TIER:

- Build and own predictive models and optimization algorithms to help TIER meet the needs of our users
- (…)
- You've always got the business and user needs in mind while you're developing. (TIER 2020a, GER)

This extract first presents data scientists as *technical* need-recognizers, and subsequently as *socially skilled* need-recognizers. The job ad is one of many that link a hybrid skill set to the question of needs. By articulating the importance of having 'business and user needs in mind', the ad implies ethical and reflective tendencies. The implicit message formulated through this discursivation is that other professionals – such as supposedly nerdy computer scientists (Kendall, 1999) – do not always have business and user needs in mind in their digital work. Data scientists are discursively associated and attributed with more social capital. Digital capitalism and the task jurisdiction of digital need discovery is thought to require a novel kind of work subject who demonstrates both need-encoding skills and social need-intelligence: a symbols person and 'shepherd' with technical and communicative capacities.

Data scientists as data exploiters and data ethicists

The third hybrid subjectivation of data scientists once again supports their positioning as professional need-discoverers. This subjectivation is based on the interpretation pattern of data scientists as *data exploiters*, on the one hand, and *data ethicists* on the other. The construction of data scientists as data exploiters manifests in the concept of 'data mining', a metaphor that is repeated throughout economic and not rarely also in academic materials to describe a core task of data scientists, namely to '[u]se data mining (…) and improve our machine learning engine' (Amazon 2019b, USA). This phrase lyrically highlights the commonalities between traditional miners, data scientists, and their end-products: both occupations use techniques to filter (or 'dig up') substances from a massive entity, a mine in one case, and a digitalized data set in the other. The metaphor points not only to a shared practice, but also to a common social figure: the exploitative subject. Data is implicitly referred to not as sensitive personal information but rather as a raw, naturally-occurring commodity – the 'new oil' as it has been dubbed. And just as the mining industry is accused of insensitivity towards nature, so too can data science be understood as producing work subjects who 'exploit' or 'extract' value from a resource, as in the examples below:

> As we operate a transaction-rich digital business model, data is at the heart of all our activities and decisions. To further exploit our data assets and become even more data driven we are looking for a Data Scientist (Fliit Holding 2020, GER)

> In our Industrial Data Science program students will learn the necessary skills to extract information out of rich data sources and to formulate recommendation based on those insights that allow companies to work more efficiently. (TU Braunschweig 2020)

The classification of data scientists as miners implies a practice of value creation with little concern for collateral damage. It connects, in this sense, with Facebook's infamous motto: 'move fast and break things'. While the term 'exploit' is used less frequently than the term 'extract' in both job ads and study programs, the meaning is quite similar: data scientists are constructed as work subjects with the ability to exploit/extract data for knowledge, which serves the purpose of primarily economic goals across both economic and academic fields. Even if the subjectivation of data scientists as data exploiters is more pronounced in the economic field, the academic field shows a surprisingly capitalist subjectivation of data scientists, given that academia traditionally presents itself as pursing different economies of worth. The prevailing of 'academic capitalism' (Slaughter and Rhoades, 1997) points not only to a consolidated neoliberal state of the field of power, where an intrusion of the logics of the economic field into the now less autonomous academic field has taken place (Bourdieu, 2000), but also reveals an important inter-field relation that undergirds the rise of digital capitalism. Contemporary digital capitalism must be understood as a system of relations that requires the labour and legitimacy not just of the economic but also of the academic fields.

However, another interpretation pattern runs counter to the figuration of data scientists as data exploiters: data scientists are also framed as *data ethicists*. As Passi and Sengers have shown, data science is not only directed towards business considerations (2020: 8). Within the academic and economic discourse, there is a parallel

discursivation of data scientists as more than capitalist work subjects. The term 'scientist' in combination with 'data' already hints at this. A framing of data science work as normative activity co-exists with the imperative for data scientists to move fast and exploit data to increase economic productivity. We can interpret these two subjectivation patterns as defining a tension between professional values, on one side, and commercial interests on the other. This configuration is not rare within occupational and professional fields (see Bourdieu, 1996a on the professional field of artists).

The ethical dimension of data science work varies between fields and national contexts. Within the economic field, *world improvement* is considered to be the central ethical dimension of data science practice. The nutrition company hundred, for instance, advertises for data scientists that want to join their 'health journey':

> [We are] a tech company in the online nutrition industry aiming to simplify the customer health journey by providing users with tailored monthly nutrition packs (…) Our data science team is in a special position to tackle problems across digital product (web, recommendations), operations research (inventory), and general behavioral prediction (probabilistic models, time series analysis). The data around our e-commerce products will interact richly with granular measures of users' health and wellness, a domain where machine learning is only recently starting to gain traction. (Hundred 2020, GER)

This ethical sense connects not only to 'solutionism' (Morozov, 2013) and 'eschatology' (Geiger, 2019: 171) but also to the dimension of needs. The mantra is that the work of data science can do good by *discovering the granular needs* of individuals/customers. For data scientists, personalization and customization appear as professional ethical convictions, a client-orientation that can be considered a typical ethos of newer 'corporate professionals' (Muzio et al., 2011: 451). On this model, data is understood to allow for an economy that strips the production system of standardization which is at odds with the values of an individualistic society.

In the academic field, taking the German case, study programs link the ethical implications of data science primarily to the dimension of data privacy. Universities present their study programs as providing preparation for data scientists to be reflective around issues of individual data rights. Especially at higher-ranked universities, specific mandatory courses exist to teach professional conduct around data: 'Students are introduced to the technical, legal, and ethical issues of data security, especially when dealing with personal data or when planning experiments in Data Science' (LMU Munich 2020). In the US-American academic field, data ethics also primarily play a role at elite universities. However, the topic is discussed from a different angle. Here, the ethical subjectivation of data scientists consists of an imperative to consider any potentially discriminatory impacts of data work. Universities in the United States tend not to thematize possible violations of individual data rights, but are instead much more concerned with practices of discrimination: 'pressing issues include: the ways in which using data can subtly exacerbate existing systemic prejudices, such as through implicit algorithmic bias' (Georgetown University 2019). The academic field in the United States positions data scientists as reflective work subjects who will consider the impacts of data work on marginalized social groups. While reflections on privacy issues and discriminatory practices can certainly overlap, they are not equivalent. The German academic field is grounded in a traditional liberal ethic while the US academic field establishes an ethic for data science that is closer to what can be considered 'identity politics' (Fraser, 2009).

Despite cross-field and national differences, the ideal data scientist is imagined as a work subject who can synthesize different expectations, which can be interpreted as a kind of 'hybrid professionalism' (Noordegraaf, 2007). At several points, the materials attribute data scientists' professional edge to their capacity to combine skills and ethics that traditionally belong to different occupational groups. Once again, hybridity is represented as the key required to unlock the proto-jurisdiction of digital need-discovery:

> You can dig deeper than most people, and you enjoy puzzles and pondering the big questions. You relish the challenge of diving into a massive amount of information and surfacing with usable and actionable insights. (…) Innovate using reflection and inquiry practices with data science to establish what they know, reveal what they don't know, and understand how they know what they don't know (University of Virginia 2019).

Thus, contrary to Zuboff's claim, surveillance capitalism is not morally agnostic (Zuboff, 2019: 381). Data scientists are constructed as work subjects who can improve need-recognition by means of an extractive ethos alongside a capacity for reflection and a passion for new knowledge. Due to their ability to decode and mine the hidden needs of customers and businesses, data scientists are presented as discoverers. In a sense, they almost come across as need-discovering psychoanalysts – professionals capable of unearthing issues that structure other actors' behaviour, but which they have been unaware of.

Discussion

Functionalists traditionally argue that objective societal needs drive professional change (Parsons, 1939). Going back to Durkheim – whose doctoral research aimed to 'investigate the function of the division of labor, that is,

the social need to which it corresponds' (Durkheim, 1997: 6) – functionalists conceptualize needs as objective requirements of social organisms. Under this rubric, the institutionalization of professions is understood as mirroring an objective need that modern, increasingly differentiated societies have. I have argued, using field and discourse theory, that it is not the functional necessities of modern societies, but rather power struggles and self-dynamic discourse that are responsible for changes in the system of professions. The need for digital capitalism for data scientists is a conflictive discursive construct. I have shown how the emergence of data science must be explored in the context of academic and economic fields as social spaces that have high symbolic capital, and I have reconstructed the discursive construction of data scientists as hybrid work subjects capable of discovering needs. Thus, the constructed functional need for digital need-discovery, rather than an objective functional need, accounts for the rise of data science.

In particular, the classification of data scientists as hybrids was interpreted as a discursive practice of distinction. While hybridity has been invoked as a subject culture in liquid modernity, here I argue that the economic and academic fields enact a discursive construction of data scientists as work subjects with conflictive skills sets and habitual dispositions, which amounts to a *cultivation of hybridity*. A central implicit knowledge order of this discourse is that data scientists are better suited for the digital data work jurisdiction than adjacent occupations and professions, such as computer scientists or statisticians, because they are not one-dimensional (Kendall, 1999). From this perspective, hybridity and plurality act not a general characteristic of modern subjects (e.g. Lahire, 2011) but rather as a specific form of cultural capital that is unevenly distributed, thus serving as a resource for power struggles. We can interpret that the constructed hybridity allows data scientists to integrate different imaginations of their role and to generally appear to be adaptive subjects – a key skill given the orientation of digital capitalism towards 'innovation' and 'disruption'. This finding leads to a better understanding of changes not only in the 'system of professions' but also in big data practices and digital capitalism in general. Since science and technology scholars have demonstrated that the cultural repertoires of tech workers find their way into the products they create (Wajcman, 2018), we can assume that the hybrid subjectivation of data scientists will impact the data products they generate. 'Classification situations' (Fourcade and Healy, 2013) are shaped by the subjectivating classifications of data scientists. Even though discourse analysis cannot show how the external construction of data scientists manifests in their bodily selves and (material) practices (ethnographic research is necessary to examine this), to some extent, they must position themselves in relation to the institutional imaginations of them. The subjectivity of data scientists thus cannot be reduced to a matter of functional necessities but must instead be considered in the context of self-dynamic discourse and field struggles.

The second main finding of this article, the discursive fact that the emergence of data science is linked to need-discovery, both reveals a second subjectivation pattern and points to how societal imaginations about data shape the role of data scientists. As demonstrated, data scientists are constructed as professionals who can unearth the otherwise unknown needs of customers and businesses through big data analytics. Beneath the subjectivation practices of the academic and economic fields lies an understanding of data as a resource that holds untapped potential for economic growth. Data is thought of as a treasure that allows for the discovery of behavioural patterns, which then enable the 'personalization' and 'customization' of products as well as the 'optimization' of business processes. The moral underpinning of big data quantification manifests in a refiguring of liberal individuality through the semantic of needs.[8]

Conclusion

I have argued that digital capitalism and issues of quantification are too often analysed without systematic concern for the growing white-collar workforce of data scientists. Based on my analysis of the data science discourse in the academic and economic fields, I have shown that the professionalization of data scientists is fuelled by their external construction as hybrid work subjects who are capable of digitally discovering needs. This hybrid construction entails three subjectivation patterns that traditionally characterize distinct social figures. These three conflictive social knowledge relationships are: (1) Data scientists are classified as both generalists and specialists; (2) Data scientists are classified as technicians and as non-nerd communicators; (3) Data scientists are classified as data exploiters, on the one hand, and data ethicists on the other. This two-dimensional subject culture is used to distinguish data scientists from competing professions often considered more one-dimensional, such as computer scientists (see Kendall, 1999). Hybridity was interpreted as a form of cultural capital that allows data scientists to appear dynamic within a digital capitalist system that is oriented towards change and 'disruption'. Furthermore, the hybrid construction is linked to the subjectivation pattern of data scientists as discoverers of needs. Data scientists are interpellated by the academic and economic field as work subjects who can unearth needs by generating and analysing large data sets. This culture of subjectivation was interpreted against the backdrop of declining growth rates in Western capitalism. It seems the professional capacity to identify new areas for economic growth has gained relevance in the labour market. There are, however, notable differences across the field-specific constructions of data scientists in the United States and Germany. The academic institutionalization of data science is further along in the United States, which

manifests in a small but growing number of autonomous data science schools and faculty. This development holds the potential to create tension between the framing of data science as a meta-discipline without a 'home turf' and its organizational institutionalization. Furthermore, US academic institutions have included communication training in their curricula for data scientists on a much broader level (even if mainly restricted to elite universities) than German institutions. In Germany, on the other hand, data science is often not only linked to the digital tech economy but also to traditional industrial companies – a difference that can be attributed to the greater relevance of manufacturing in Germany. Lastly, cross-national differences also exist with regard to ethics. In the German context, data scientists are trained to consider individual privacy issues, while US-Americans put greater emphasis on 'identity politics' (Fraser, 2009) and considering group-based discriminatory practices.

In essence, though, the discursive constructions of data scientists show more commonalities than differences across both national contexts, in which data scientists are systematically classified as hybrid discoverers of needs. The construction of data scientists in non-Western contexts would be an interesting topic for further research. As delineated, it is also necessary to complement this study with an inquiry into other professionals, such as statisticians, that also compete over the digital data work jurisdiction. Another issue that lies beyond the scope of this article is the question of how the discourse about data science actually manifests in the practices of data scientists. In recurrence to Critical Data Studies (Iliadis and Russo, 2016) and Science and Technology Studies (Wajcman, 2018), this line of interest could be further pursued through an analysis that focuses on data science as constituted within wider data assemblages. Existing ethnographic inquiries into the practices of data scientists have contributed important insights along these lines (Knox and Nafus, 2018; Mützel et al., 2018; Passi and Sengers, 2020; Slota et al., 2020). And since digital capitalism forms the breeding ground not only for data scientists but also for a number of other so-called 'tech workers' (Tarnoff and Weigel, 2020), there is also a need to explore adjacent professional fields including 'UX-Designers' or 'AI Engineers'. A final issue that should be raised is the question of which individuals are habitually best equipped to perform hybridity, which is considered so vital in the data science profession for the fulfilment of need-discovery. Assuming that professional expertise is not located within individuals, but in relationships between them, this desideratum would return us to questions of social class (Savage et al., 2013).

Acknowledgements

Thanks for helpful comments are due to Anders Blok, Philipp Brandt, Florian Butollo, Franziska Cooiman, Florian Eyert, Jennifer Gabrys, Martin Krzywdzinski, Steffen Mau, Gillian Moore, Sebastian Raza, Andreas Reckwitz, Philipp Staab, Teresa Völker, Jan Wetzel as well as the four anonymous reviewers.

Declaration of conflicting interests

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

Funding

The author received financial support from the Department of Sociology at the University of Cambridge, the Foundation of the German Economy, and the Kursbuch Kulturstiftung.

ORCID iD

Robert Dorschel https://orcid.org/0000-0002-8839-2752

Notes

1. Barry and Charpentier (2020) have explored the 'personalization promise' of big data. Zuboff has discussed mass personalization with regard to the entangled hope for new economic growth (Zuboff, 2019: 254).
2. Of course, though, how this interpellated hybrid subjectivity is embodied by data scientists in 'non-discursive practice' is an empirical question for further research. Research combining ethnographic analysis and discourse analysis could provide insights into how discourse is performed or resisted by actors.
3. Abbott later developed the concept of 'linked ecologies' to account for this gap (Abbott, 2005). But Abbott's theory lacks concepts, such as habitus or capital, that would allow for a micro- and macro-theoretical understanding of power struggles.
4. Foucault and Bourdieu are both ultimately interested in how power generates the cultural schemas that are incorporated by individuals. However, Bourdieu's concept of the habitus is tied to class and fields (Bourdieu, 1990: 59), while Foucault's notion of subjectivation highlights the entanglement of individuating practices with discourses, dispositifs and technologies of the self (Foucault, 1982).
5. Again, this is not to say that discourse is the sole force that has contributed to the professionalization of data science. A Foucauldian approach to any social phenomenon acknowledges its relations to discursive as well as non-discursive practices. Thus, the rise of data science is certainly also tied to economic, technological and organizational changes. Discourse cannot, however, be regarded as a superstructural dimension on top of these changes.
6. I relied on the rankings of the 'Times Higher Education World University Ranking 2018' for the sampling. Certainly, rankings are controversial. However, they matter in the sense that they constitute an objectified reputation system which actors increasingly turn to for orientation.
7. I considered companies to be startups if they were less than ten years old and were oriented towards innovation and fast growth.
8. Hitherto, social theory has not developed a systematic constructivist concept of needs.

References

Abbott A (1988) *The System of Professions*. Chicago: University of Chicago Press.

Abbott A (2005) Linked ecologies: States and universities as environments for professions. *Sociological Theory* 23(3): 245–274.

Barry L and Charpentier A (2020) Personalization as a promise: Can big data change the practice of insurance? *Big Data & Society* 7(1): 1–7. SAGE Publications Ltd.

Blok A, Lindstrøm MD, Meilvang ML, et al. (2019) Ecologies of boundaries: Modes of boundary work in professional proto-jurisdictions. *Symbolic Interaction* 42(4): 588–617.

Bourdieu P (1988) *Homo Academicus*. Stanford: Stanford University Press.

Bourdieu P (1990) *The Logic of Practice*. Stanford: Stanford University Press.

Bourdieu P (1991) *Language and Symbolic Power*. Massachusetts: Harvard University Press.

Bourdieu P (1996a) *The Rules of Art: Genesis and Structure of the Literary Field*. Stanford: Stanford University Press.

Bourdieu P (1996b) *The State Nobility: Elite Schools in the Field of Power*. Cambridge: Polity Press.

Bourdieu P (2000) *Backfire: Against the Tyranny of the Market*. Cambridge, UK: Polity Press.

Bourdieu P (2005) Principles of an economic anthropology. In: *The Handbook of Economic Sociology*. Princeton, NJ: Princeton University Press, 75–89.

Bourdieu P and Wacquant LJD (1992) *An Invitation to Reflexive Sociology*. Chicago: University of Chicago Press.

Brandt P (2016) *The Emergence of the Data Science Profession*. PhD Dissertation.

Braverman H (1998) *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*. New York: New York University Press.

Callon M (1984) Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. *The Sociological Review* 32(1): 196–233.

Couldry N and Mejias UA (2019) *The Costs of Connection: How Data is Colonizing Human Life and Appropriating it for Capitalism*. Stanford: Stanford University Press.

Crouch C (2019) *Will the Gig Economy Prevail?* Cambridge: Polity.

Davenport TH and Patil D (2012) Data scientist. *Harvard Business Review* 90(5): 70–76.

Durkheim E (1997) *The Division of Labor in Society*. New York: Free Press.

Foucault M (1982) The subject and power. *Critical Inquiry* 8(4): 777–795.

Foucault M (2002) *The Archaeology of Knowledge*. New York: Routledge.

Foucault M (2003) *The Birth of Clinique: An Archaeology of Medical Perception*. London: Routledge.

Foucault M (2005) *The Order of Things*. London: Routledge.

Fourcade M and Healy K (2013) Classification situations: Life-chances in the neoliberal era. *Accounting, Organizations and Society* 38(8): 559–572.

Fraser N (2009) Social justice in the age of identity politics. In: *Geographic Thought: A Praxis Perspective*. New York: Routledge, 72–91.

Fuchs C (2013) *Digital Labour and Karl Marx*. New York: Routledge.

Gabrys J (2011) *Digital Rubbish: A Natural History of Electronics*. Ann Arbor: University of Michigan Press.

Gabrys J (2016) *Program Earth: Environmental Sensing Technology and the Making of a Computational Planet*. Minneapolis: University of Minnesota Press.

Geiger S (2019) Theorizing disruption through an eschatological lens: Silicon Valley, inevitability and the end of history. *Journal of Cultural Economy* 13(2): 169–184.

Goldstein J (1984) Foucault among the sociologists: The "disciplines" and the history of the professions. *History and Theory* 23(2): 170–192.

González-Bailón S (2017) *Decoding the Social World: Data Science and the Unintended Consequences of Communication*. Cambridge: MIT Press.

Graham M, Hjorth I and Lehdonvirta V (2017) Digital labour and development: Impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer: European Review of Labour and Research* 23(2): 135–162.

Hall PA and Soskice D (2001) *Varieties of Capitalism. The Institutional Foundations of Comparative Advantage*. Oxford: Oxford University Press.

Hammerbacher J (2009) Information platforms and the rise of the data scientist In: *Beautiful Data: The Stories Behind Elegant Data Solutions*. Sebastopol: O'Reilly Media, 82–94.

Iliadis A and Russo F (2016) Critical data studies: An introduction. *Big Data & Society* 3(2): 1–7.

Ingram N and Allen K (2019) 'Talent-spotting' or 'social magic'? Inequality, cultural sorting and constructions of the ideal graduate in elite professions. *The Sociological Review* 67(3): 723–740.

Keller R (2011) The sociology of knowledge approach to discourse (SKAD). *Human Studies* 34(1): 43.

Kendall L (1999) Nerd nation: Images of nerds in US popular culture. *International Journal of Cultural Studies* 2(2): 260–283.

Knox H and Nafus D (2018) *Ethnography for a Data-Saturated World*. Manchester: Manchester University Press.

Kotras B (2020) Mass personalization: Predictive marketing algorithms and the reshaping of consumer knowledge. *Big Data & Society* 7(2): 1–14.

Lahire B (2011) *The Plural Actor*. Cambridge: Polity.

Lehdonvirta V (2018) Flexibility in the gig economy: Managing time on three online piecework platforms. *New Technology, Work and Employment* 33(1): 13–29.

Mackenzie A (2018) Personalization and probabilities: Impersonal propensities in online grocery shopping. *Big Data & Society* 5(1): 1–15.

Mau S (2019) *The Metric Society: On the Quantification of the Social*. Cambridge: Polity.

Metcalf J and Crawford K (2016) Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Society* 3(1): 1–14.

Morozov E (2013) *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: Public Affairs.

Mützel S, Saner P and Unternährer M (2018) Schöne Daten! Konstruktion und Verarbeitung von digitalen Daten. In: *Datengesellschaft. Über Die Datafizierung Des Sozialen*. Bielefeld: Transcript, 111–132.

Muzio D, Hodgson D, Faulconbridge J, et al. (2011) Towards corporate professionalization: The case of project management, management consultancy and executive search. *Current Sociology* 59(4): 443–464.

Noordegraaf M (2007) From "pure" to "hybrid" professionalism: Present-day professionalism in ambiguous public domains. *Administration & Society* 39(6): 761–785.

O'Neil C and Schutt R (2013) *Doing Data Science: Straight Talk from the Frontline*. Sebastopol: O'Reilly and Associates.

Parsons T (1939) The professions and social structure. *Social Forces* 17(4): 457–467.

Passi S and Sengers P (2020) Making data science systems work. *Big Data & Society* 7(2): 1–13.

Riesman D, Glazer N and Denney R (2001) *The Lonely Crowd: A Study of the Changing American Character*. New Haven: Yale University Press.

Sadowski J (2019) When data is capital: Datafication, accumulation, and extraction. *Big Data & Society* 6(1): 1–12.

Saner P (2019) Envisioning higher education: How imagining the future shapes the implementation of a new field in higher education. *Swiss Journal of Sociology* 45(3): 359–381.

Savage M, Devine F, Cunningham N, et al. (2013) A new model of social class? Findings from the BBC's Great British Class survey experiment. *Sociology* 47(2): 219–250.

Sennett R (2008) *The Craftsman*. New Haven: Yale University Press.

Shestakofsky B (2017) Working algorithms: Software automation and the future of work. *Work and Occupations* 44(4): 376–423.

Slaughter S and Rhoades G (1997) *Academic Capitalism and the New Economy: Markets, State, and Higher Education*. Baltimore: JHU Press.

Slota SC, Hoffman AS, Ribes D, et al. (2020) Prospecting (in) the data sciences. *Big Data & Society* 7(1): 1–12.

Staab P and Nachtwey O (2016) Market and labour control in digital capitalism. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 14(2): 457–474.

Tarnoff B and Weigel M (2020) *Voices of the Valley. Tech Workers Talk about What They Do – and How They Do It*. New York: Farrer, Straus Giroux.

UNCTAD (2019) United Nations Digital Economy Report. Available at: https://unctad.org/webflyer/digital-economy-report-2019

Wajcman J (2018) How Silicon Valley sets time. *New Media & Society* 21(6): 1272–1289.

Zuboff S (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.