8

Statistician Subjects: Differentiating and Defending

Francisca Grommé, Baki Cakici, and Ville Takala

Introduction[1]

Who are the professional subjects of data practices? How are their skills, capacities, mindsets, and ethical positions shaped in relation to data practices? While the previous chapter explored how data practices subjectify people and how they are categorised, here we turn to consider how the statistician subject is being shaped, and the profession of national statistician repositioned, through what we refer to as 'professionalising practices'. The chapter develops this understanding by returning to technological changes described in the Preface: how digital technologies such as the internet, mobile devices, and big data are both challenging traditional methods of producing official statistics while at the same time offering possibilities to innovate the production of statistical knowledge about populations. We consider how data scientists are leading the development of such methods, especially those that involve big data.

As this chapter sets out, it is through the valuing and performing of data practices that engage with big data and related analytical techniques (which we will herewith refer to simply as big data) that the statistician subject is being shaped and the profession of national statistician repositioned. As expressed by leading national statisticians at conferences and in policy

papers, to meet the challenges and realise possibilities of big data requires more than simply changing data practices. It also requires that statisticians develop skills and knowledge not typically deployed in the production of official statistics, for example, analytic techniques such as machine learning and predictive modelling. Yet, what they also acknowledge is that skills alone are insufficient. 'Cultural change' is also necessary as advocated by proponents of the uptake of big data. As one speaker at a 2015 international seminar about the future of official statistics noted: 'Machine learning is the future. Big data cannot be processed by hand. Therefore, the current culture is a liability.' This speaker specifically referred to routine data practices, where the production of statistics often requires professional human judgement. However, as they observed, large volumes of data require automated forms of data processing which involve less human intervention. In their view, this is one way that the current culture of official statistics is being challenged.

Generally, cultural change has come to refer to a broad, fuzzy set of organisational, practical, and other desired changes in the profession of statistician.[2] This includes their skills and ways of thinking and a direction of change that are in part modelled on various understandings of the 'entrepreneurial mentality' of data scientists working in the technology sector. For instance, in a report on the value of official statistics, a UNECE task force provided an inventory of corporate practices, which included those of Apple, Amazon, and Google. Even though the report notes that statisticians 'have considerable comparative advantages' (UNECE, 2018: 10) to meet the needs of an information age, it also states the following:

> But competing information providers [e.g., Google] have advantages, too. Sometimes, they will have resources available to them which dwarf those available to most NSOs [national statistical

> organisations]. They may also have cultures which allow them to take up new technologies and methodologies more quickly than traditionally has been the case in the official statistics community. They may also have cultures, driven by commercial necessity, which make them more responsive to customer needs (UNECE, 2018: 10).

What the quote epitomises is that private sector data providers are emerging competitors of NSIs. Data scientists in the private sector have the expertise and skills to 'take up new technologies and methodologies' required, and which are valued in relation to those of national statisticians.

We approach these questions of skills, expertise, mentalities, and cultures required to take up big data by considering them as objects of valuation and struggle within what we have previously conceived of as the transnational field of statistics (Scheel et al., 2016). As we will elaborate below, it is through struggles that the faction of national statistician is competing with other professions over the relative valuation of cultural capital and habitus required to work with big data. Such competition is occurring mostly in relation to data science and its professional subject, the data scientist. Yet, what constitutes data science or a data scientist is not universally agreed nor stable[3] (as is the case for other scientific disciplines and professions). Contemporary definitions of data science and data scientists are closely associated with big data, a term that became mainstream around 2011. In addition, the relations between official statistics and data science are framed in different ways. Whereas many frame them as competitive, some statisticians publicly speak out against a division between data science and statistics, arguing that statistics are at the core of data science and that the volume of data does not change that fact (cf. Meulman, 2016).

Professionalising practices are part of such struggles. In some situations, this involves recognising forms of cultural capital and cultivating a habitus aligned with conceptions of the faction of 'data scientist', while in others it involves defending the faction of national statistician;[4] both situations are the object of the analyses that follow. Our aim is twofold: first, to understand *how* such change is being pursued through professionalising practices. Considering insights from STS and related fields, we do this by analysing how skills, capacities, mindsets, and ethical positions are valued discursively through job interviews, but also performed through material-semiotic practices such as data camps. Our second aim is to consider how such professionalising practices of national statisticians involve a tension between entrepreneurial and public service skills and habitus.

In what follows, we first elaborate our conceptualisation of the transnational field of statistics. Next, we empirically examine the shaping of the statistician subject and the repositioning of the faction of national statistician by analysing three professionalising practices: recruitment job interviews in the UK; a brainstorming workshop and data camp modelled after hackathons at Statistics Netherlands (SN); and presentations by statisticians at Eurostat and UNECE conferences. In the conclusion we highlight how thinking about professionalising practices is important to understand how data practices do not simply involve struggles over methods of producing statistics. They also involve professional struggles over the skills and habitus that are valued and cultivated. That is, to advocate the valuing of a particular data practice also involves recognising the required skills and habitus to perform them and in turn the relative advantages of professionals who possess them. We then suggest that data practices are bound up with professionalising practices and need to be considered together

to investigate the politics of method and the production of official statistics.

Shaping the Professional Subject in Relation to Big Data

As noted, we conceive of professionalising practices as part of struggles over the legitimacy of methods and their related data practices in the production of official statistics.[5] Such struggles are situated in, and help to shape, what can be understood as a transnational field of statistics. It is through specific practices that actors from competing professions attempt to advance or defend their relative positions within a field (Bourdieu, 1989). The field of statistics comprises differently positioned professions such as statisticians, demographers, domain specialists, academics, policy makers, and other users of statistics (cf. Scheel et al., 2016). While statisticians have long occupied a dominant position, data scientists are an emerging faction challenging this dominance within the field. For data scientists, at stake is recognition of big data and related analytical methods as legitimate and authoritative and, in turn, the cultural and symbolic capital that this will confer. For statisticians, their stakes are to protect and advance their authority and position in relation to each other and this faction. Through this understanding of the field we conceive of these stakes as a politics of method, as it provides a way to analyse 'the emergence of new kinds of practices' (Bigo, 2011: 240–241). In brief, the transnational field of statistics involves struggles over data and methodological innovations and authority in the production of official statistics. We understand this as 'a messy, competitive context [in which] the roles of different kinds of intellectuals, technical experts and social groups are at stake' (Savage, 2010: 237).

This competition does not only involve claims about data and the positioning of the existing methods of official statistics vis-à-vis those of data science; it also involves establishing the statistician subject as a trustworthy and competent professional. Here, we consider how the statistician subject is shaped by their socialisation within the field of statistics. This dynamic works in two directions: professional subjects are both shaped by the field and come to shape it through their practices. The factions of national statistician and data scientist are distinguished from each other by the valuation and appropriation of certain forms of cultural capital over others. Cultural capital includes skills, but also the knowledge of what to value and what professional ethics to support. This positioning is not entirely a matter of conscious choice. Instead, orienting towards a position in a field 'functions below the level of consciousness and language and beyond the scrutiny or control of the will' (Bourdieu, 1984: 466). Statistician subjects thus take up positions both as a result of the valuations of some skills, normative inclinations, and dispositions above others (for instance, in textbooks or by authority figures), and come to also embody these. The particular combination of skills, habits, normative inclinations, and so on that subjects come to embody constitute what Bourdieu refers to as 'habitus' ('a system of dispositions').[6]

Two examples well illustrate how the introduction of new methods to know populations can affect the formation of the professional subject. In the field of statistics, Savage shows how a shift from the 'gentlemanly social scientist' to a professional with a 'technical orientation' took place in the 20th century through the invention of the sample survey (2010). He argues that for the sample survey to be recognised as legitimate, differently positioned actors within the field needed to be convinced of the trustworthiness and validity of interview. The statistical

technique of random sampling was advanced to support the reliability of data as it reduced selection bias and led to more representative statistics. The consolidation of the interview method in turn supported the rise of social science and technically oriented researchers as part of the state apparatus. The second example is from the field of border security. Bigo examines how data collection technologies were relevant in distinguishing different professional positions within the field.[7] For instance, database analysts formed their professional positions around the authority of 'smart technologies' (such as predictive software). Bigo argues that dispositions within the field were 'activated – or not, as the case may be – by the use of specific technologies, and [dispositions] determine the capacity to restrain the deployment of these technologies, to modulate them' (Bigo, 2014: 210). Consequently, professional dispositions are not determined but can be activated, shaped, and reinforced by technologies deployed to know targe populations.

STS studies have pointed out that professional subjects are shaped, and positions valued and inhabited, through not only discursive but also material-semiotic practices. In their study of the rise of the experimental technique during the English Restoration, Shapin and Schaffer (1985) show how the social technology of prescribing 'modesty' was a key attribute of the formation of the experimental scientist. As Haraway (1997) later pointed out, such technologies also positioned scientist subjects as essentially male. Following these and other studies (see, for instance, Latour, 1993), Ruppert and Scheel demonstrated how the dynamics that arise when new methods are introduced within the field of statistics cannot be reduced to discursive claims:

material-semiotic practices like demonstrations that seek to legitimize innovations in methods and data as official. In this way, we

> underscore that the politics of method are not reducible to a competition between human actors who can put forward the best argument in the most compelling manner. Rather, the politics of method requires a symmetrical analysis that accounts for how different kinds of digital devices are mobilized in struggles over methodological innovations in the production and legitimation of official statistics (Ruppert and Scheel, 2019: 3–4).

In this chapter we adopt this focus on the relevance of both discursive and material-semiotic practices. We suggest that professionalising practices 'make explicit' the cultural capital and habitus involved in the formation of professional subjects. As Muniesa and Linhardt explain, 'making explicit' involves 'the actualization of the virtual' and 'about expressing something, provoking it in variable, conflicting, unanticipated manners, putting it to the test of becoming an actual configuration, an actual event' (2011: 546). Making things explicit does not unfold without problems, hesitations, or tensions. Rather, sensibilities are made visible and can then be put up for consideration, debate, or negotiation whether they arise in job interviews, brainstorming workshops, or conference presentations.

In the empirical sections below, we examine how professionalising practices make explicit tensions and congruities between the cultivation of entrepreneurial and public service skills and habitus. Our analysis of an entrepreneurial habitus draws on two ethnographic studies. The first concerns Irani's study of professional designers in India (2019). Characterising the tech sector, Irani shows that an entrepreneurial disposition includes a strong belief in technological innovation as the prime locus for societal change (instead of, for instance, poverty alleviation policies). Further aspects of an entrepreneurial disposition are a sense of optimism and urgency to

accomplish innovation and a strong belief in collaboration as the key to solving complex issues and problems. In addition, it includes the practice of experimentalism, in the sense that work is not always aimed at producing immediate tangible results. This notion of experimentalism is not only embraced to support learning through trial and error, it is also embraced because it allows for suggesting or hinting at future potential and value:

> But it is not tangible productivity, but what anthropologist Kaushik Sunder Rajan characterises as the 'felt possibility of future productivity or profit' (2006, 18). They produce and respond to vision, hope, and hype as they pursue speculative capital and investment; they promise not only financial value but also social value and legitimation for socially responsible funders and investors (Friedner 2015) (Irani, 2019: 16).

Even though statisticians are less affected by technology hype cycles and the pressures of external investors, this aspect of an entrepreneurial disposition may be relevant as statisticians do need to attract internal and external support and funding for new ideas.

Mackenzie's (2013) study of the practices of data scientists connects this disposition to data practices that involve the use of machine learning. The adoption of predictive analytics in these practices, he demonstrates, is part of a habitus that embraces probabilistic outcomes, likelihoods, and the optimisation of models, rather than their verification (as in 'traditional' statistics). This logic of optimisation and prediction can be applied to problems in a wide range of social domains, which is a key feature of entrepreneurialism. Finally, Mackenzie shows that an entrepreneurial ethos is further internalised by data scientists through competitions and

hackathons that involve a rhetoric of addressing them as 'wonderful people' (2013, 394): a highly desirable group equipped with a unique combination of skills to address the social challenges of our times.

What these studies offer is that an entrepreneurial habitus can be at work in several connected ways: in the development of skills and sensitivities to identify potential 'social problems' (and thereby potential markets), as well as in the appreciation and internalisation of a particular set of methods and related sensitivities. We will explore how valuations of future potential are relevant for the shaping of the statistician subject but are in tension with those of public service, a tension that is made explicit in professional practices.

Recruiting Data Scientists

Looking for Data Scientists

The first professionalising practice that we explore is government recruitment interviews for data scientists in the UK. Through the analysis of job descriptions and interviews we consider how valuations of 'data scientists' and the professional skills necessary for working with big data are made explicit. Specifically, we highlight how data scientists are differentiated from national statisticians and valued in relation to their 'future potential'. At the same time, we show how national statisticians are differentiated and valued in relation to their public service skills and dispositions.

In 2015, a recruitment committee interviewed applicants for data scientist posts distributed across several government departments. The committee included a statistician of the NSI, and two other civil servants, one from the human resources department of the NSI, and one from another government

agency. Each interview lasted from 45 minutes to one hour, was situated in a small room, and involved a question-and-answer exchange between the recruitment committee and the applicant. The applicants were expected to demonstrate how their previous experience and knowledge were compatible with the role of the data scientist. Meanwhile, the committee needed to reach a consensus on whether the applicants' responses fulfilled the requirements for becoming a data scientist. The applicants were also required to take a multiple-choice test in an adjacent room following their interviews, which included questions on basic statistics knowledge such as the definition of terms, probability calculations, and so on.

The job description document that advertised the position presented an ideal type of data scientist: someone with a collection of skills in programming, computing, data, and statistics. The interview committee was asked to formulate questions in relation to this description to assess the candidate's competency in different skills. They were also provided with a 'marking matrix', a document listing the categories and the grades they should use to assess the performance of the applicants during the interview. This matrix outlined the data scientist profession across two categories of questions, 'job specific' and 'competency', each with four subcategories. Job-specific categories referred to the technical skills of data scientists: 'computing' focused on programming languages; 'scripting' emphasised experience in using statistical tools such as R, SAS, SPSS; 'software' referred to big data analytics tools such NoSQL, Hadoop, Spark, and so on; and 'statistical skills' as the knowledge of traditional statistical methods, such as how to determine if a sample is representative. The competency category included references to broader skills that are applicable to all civil service positions and define a common core of skills and dispositions that civil servants are

expected to possess: collaboration, personal improvement, meeting deadlines, leadership, and communication with an emphasis on the ability to explain technical issues to non-technical audiences. Under this category, the job interviews defined the position of a data scientist but also differentiated it in relation to the national statistician by introducing public service skills and dispositions.

Of note is that the data scientists sought in the interviews were not being hired for a specific government task or practice. They could be placed in different government departments, but still expected to contribute their own skills independent of the domain. In other words, the cultural capital of the data scientist was conceived as highly convertible, allowing them to work in different domains with the same set of skills (cf. Mackenzie, 2013). However, all were expected to perform as civil servants in ways listed under the competencies category of the marking matrix.

The question-and-answer session made explicit many of the skills and values at stake in defining data scientists, but also for advancing and valuing the skills of national statisticians. To prove their potential as government data scientists, the candidates were expected to demonstrate their statistical expertise by answering questions such as 'How do you know if your result is statistically significant', or 'How did you know if your sample represented the population?' When one of the candidates provided inadequate answers to these questions, the interviewers added a note to their application during the assessment round, asking him to 'please look at the statistical techniques required [for the position]'. Statistical tools were also discussed, as most candidates brought up Matlab, SAS, SPSS, and R when asked about their experience with software. R, short for the R Project for Statistical Computing, was often emphasised as the ideal tool due to its status as open-source

software, but also because it was 'less clunky than SPSS', in the words of one candidate. The interviewers also queried the applicants' familiarity with big data through questions such as: 'What did you learn from your experiences working with big data projects?' to which one candidate replied 'Use fewer programming languages', which displayed their familiarity with a shared perception within data science of the proliferation of tools and languages. Through their answers, the candidates implied that some new technologies were used in a project for the sake of having used them, and that such uses did not belong in 'proper' data science. Consequently, not only knowledge of skills and tools were tested and demonstrated, but also preferences and subtle distinctions.

Following each interview, the committee members were required to individually assign different scores to the eight subcategories in the marking matrix based on a scale from one to seven. The evaluation also involved a multiple-choice assessment for some categories, where the interviewers were expected to tick under 'positive', 'needs development', or to leave it blank. The committee filled in their forms individually, and then discussed their answers to reach consensus on the final assessment of a candidate, which did not prove very difficult as their assessment of most categories were either the same, or very similar. During one such discussion, a committee member stated that given sufficient background such as a quantitative PhD, or prior experience in statistical programming, the applicants would be able to pick up some of the necessary skills even if they did not seem to possess them at the time of interview. In other words, they evaluated the applicants' potential to become data scientists. As explained next, some of this potential was articulated by referring not to data science skills, but to a set of skills that differentiate them from those of national statisticians.

Data Scientist As …

Recruitment processes are framed prior to interviews in application documents such as the job description, guidance for candidates, sample multiple-choice tests, and other supporting texts, as well as those submitted by applicants in the form of CVs and test answers. These documents describe the profession, and list expected skills, but the recruitment process is far from an exercise of fitting people into predefined boxes; the situated performing of the job interview also refines what it means to belong to a profession.

Who, then, are the data scientists as enacted by the job interview? They can program, acquire new technical skills quickly, have basic statistical knowledge, be familiar with the discourse of big data, be reflexive about not only the division between the highly technical and the traditional statistical, but also their own position within various government departments. They are not merely programmers or developers as they also possess statistical expertise, but they are also more than just methodologists as they do not rely on other developers to conduct their study or produce their results. The data scientists combine statistical knowledge with new forms of data analysis. At the same time, the data scientists of the job interview are not hackers. They do not solve problems through small, localised fixes. Instead, they follow specific methodologies informed by traditional statistical data practices.

In short, the job interview enacted the data scientist as possessing a set of skills and dispositions. Candidates were expected to possess cultural capital in the form of particular accumulated technical skills such as statistical analysis and programming that could be converted to advantage in the ongoing struggle to define the profession of data scientist (Halford and Savage, 2010). The candidates needed to possess

certain cultural capital such as statistical expertise and related technical skills to succeed in the recruitment process, but as the interviewers also acknowledged, the interview included an evaluation of their potential to become data scientists. That is, being a data scientist involved a process that built on cultural capital that a candidate already possessed but through the recruitment interview they needed to also perform the capacity to learn and acquire yet unknown skills. In this need to build on something, we identify their relation to the faction of national statistician. To become a data scientist involves a process of accumulating cultural capital beyond that possessed by statisticians, such as new programming languages, or familiarity with new data analysis tools as technologies change and evolve. The situated performance of the recruitment interview is where such future potential is assessed.

While recruitment interviews for data scientists valorised new skills in the data practices of government, skill alone was not sufficient. It needed to be bundled with other forms of cultural capital such as statistical knowledge as a foundation, as well as the habitus of a civil servant. However, in this specific bundle, technical skills counted for more when granting legitimacy to the performance of the data scientist candidate. When applicants argued for why different skills should be considered part of the bundle, they built on those of the profession of national statistician as a foundation while also differentiating theirs as more than just those of a national statistician. Some skills, for example familiarity with database management, a task once relegated to IT-specialists, played a much more prominent role, defining the data scientist and differentiating it from that of national statistician.

In these ways the situated performance of the recruitment interview made explicit the differentiation between the two factions in the field of statistics. What the valuations of

'future potential' and the competencies of public service skills and dispositions show is the forms of cultural capital that are stakes in the struggles for recognition in the field of statistics.

Innovation Events: Brainstorming Workshop and the Data Camp

From the Potential of a Job Candidate to the Potential of Big Data

To experiment with big data, NSIs need statisticians not only with data science skills but also 'big data sensibilities', as stated by a senior national statistician. Such sensibilities can be understood as making up a data scientist habitus: embodied cultural capital that includes tastes, habits, normative inclinations, and other knowledges and sensibilities that are not normally made explicit. Just as the skills that make up a data scientist and how they differ from or resemble those of national statisticians emerges through interviews, what constitutes the habitus of data scientists emerges through specific material-semiotic professionalising practices. We develop this by discussing our observations of two professionalising practices focused on innovation and organised by Statistics Netherlands (SN): a brainstorming workshop and a data camp. The events took place in the context of a wider debate within SN on the uptake of big data. Several introductory sessions and presentations took place before both events during which some statisticians regularly expressed their scepticism towards big data. For instance, a frequent objection to using social media data was that it is not representative of a national population, and that relevant background characteristics (age, gender) cannot be verified. In addition, using social media data would imply diverting from the international definition of a statistical

population, that is, usual residents (for instance, a Twitter population can also include tourists). As one statistician phrased it: 'this would be very dangerous.' But the events also involved a small group of statisticians within the NSI who already had an active interest in adopting big data and thus were committed to engaging in experimentation; this was the group participating in the two professionalising practices discussed in this section.

We first discuss the data camp to highlight how it fostered an entrepreneurial disposition necessary to develop future-oriented, 'risky' projects (Irani, 2019). Next, we highlight how this disposition included the capacity to work with techniques and visualisations that can demonstrate the potential of big data and data science. Rather than performing the future potential of job candidates to be data scientists, as elaborated in the previous section, the brainstorming workshop and data camp involved performing a disposition necessary to demonstrate the potential of big data. Furthermore, this disposition was not only performed and cultivated discursively, but also through material-semiotic data practices.

From Skills to Sensibilities

The aim of the brainstorming workshop was to develop ideas for public sector innovation by combining different types of data. It was organised by the NSI's innovation lab for the purpose of developing submissions to a competition organised by the Ministry of Economic Affairs. Eight people from different backgrounds and positions took part in a two-hour session, led by an NSI innovation expert. We first take a closer look at the fostering of a set of sensibilities or dispositions focused on the development of new projects.

From the outset, it was clear that to brainstorm is not only a cognitive process but also involves particular dispositions.

This included working at a fast pace and generating ideas quickly. For example, as participants (including one of the authors) gathered around a flip-chart they were advised to 'stay active' by standing up (lunch would be a stand-up lunch) and walking around. Moments for 'inward' individual reflection were limited in favour of fast and collaborative idea generation. All ideas counted; inhibitions and concerns were cast aside for the duration of the session.

Relevantly, the ideas to be generated were not randomly determined but targeted to particular goals; they needed to be future oriented, as the session leader explained: 'we need to be anticipatory, so that when it becomes relevant, we have the data … We want to know what the relevant issues will be in two years.' Following these instructions, the participants came up with a list of topics that included 'robotisation,' 'clean drinking water,' and 'what Google knows about its users.' The group leader made a point of explaining the difference. Statistics would not just be 'user-oriented,' that is, tailored to the needs of policy makers, journalists and other user groups. Rather, the focus was on future 'social problems' to be mitigated by data analysis. Examples of such problems, as the group leader explained, are clean drinking water, whereas needs refer to the statistics that users state that they require.

But how then are social problems to be identified in advance? Or, as some participants phrased it, how can we 'get the signal from society'? New analytical techniques and media were discussed at length as possible solutions: 'We could vlog [produce YouTube reports], start an online focus group or become data journalists. As long as we can get the signal.' To apply these techniques, someone else said: 'We need to be able to take risks, to experiment, for instance by monitoring social media.' Referring to Google's Project X (a secretive and high-risk research facility funded by Google), she proposed to

form a 'risk taking group' with everyone present in the brainstorming workshop and other interested employees. To this another participant responded: 'the civil servant rebels! They [the "risk taking group"] are here to sacrifice sacred cows … they experiment with data and possibilities!' The brainstorming workshop thus helped develop a particular disposition required to generate ideas and projects that anticipate future social problems.

Whereas the brainstorming workshop was a short and focused event, the data camp was more immersive. It was attended by national statisticians, students, and researchers with PhDs in computer science and related disciplines. The format loosely imitated a hackathon, and included skills training, lectures, presentations, and group work. Twenty participants and seven mentors from SN and a university stayed on a university campus for a week. The mixed NSI-university teams worked until late at night on topics, not even stopping work during the 'data dinners'. Among the projects initiated by the teams were the analysis of Twitter data to learn whether gender can be derived from profiles or statements; the use of Twitter statements to predict tourist behaviour or crowded events; the use of road sensor data to predict economic growth; and the use of citizen science data to model the development of the blooming phase of flowers over space and time.

The data camp demonstrated how the ability to articulate the potential of big data necessitated acquiring sensibilities about particular analytical techniques and their aesthetics. Three sensibilities taught at the data camp illustrated this. The first was an 'appreciation of algorithms'. In the plenary sessions following the group work, and in reports about the group project, participants mentioned the relevance of algorithms, by which they referred to the commands and codes that help them execute a wide variety of automated work: converting data

sets, classifying data for more insight and analysis, codes that extract and select relevant data, mining text, calculating values, and finally implementing analytic models. Correspondingly, in their evaluations and reports, participants emphasised the relevance of algorithms to process data and to get insight into data sets.

But such statements of relevance amounted to more than simply acquiring skills. Some participants stated that their work required an 'appreciation' of algorithms. For instance, one of the reported outcomes of an evening evaluation session was that 'algorithms love statistics' (see Figure 8.1). As automated correction and processing work also happens at NSIs, using algorithms was not new for statisticians. Yet in this instance statisticians referred to an intimacy between algorithms and statistics that helped them not only understand but also to realise the potential of large data sets to clean data so it can be analysed early on in the production process. However, participants emphasised in the plenary sessions that algorithms did not necessarily make data processing and analysis quick and simple tasks – they required patience. A data science habitus thus included an appreciation for algorithms paired with the virtue of patience to realise the potential of data (see Figure 8.1).

The second sensibility was a preference for a particular visual aesthetic; two lectures about visualisation during the camp are especially instructive. The first was given by the CEO of an NGO working according to the principle of what he referred to as 'objects of concern' (drawing on the work of Bruno Latour).[8] The speaker argued that this can require increasing the visibility of a local phenomenon (like deforestation) on a map, in order to draw attention to it. 'You have to take a position', the organisation's CEO stated, 'not exaggerating is

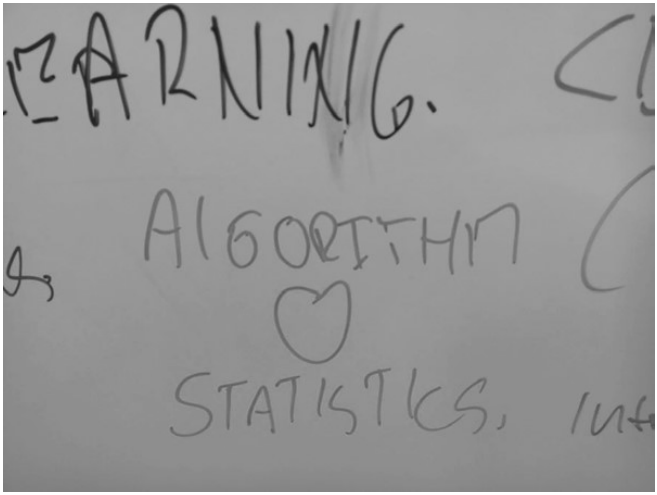

Figure 8.1 The Love of Algorithms
Source: Photo of the Data Camp's Whiteboard after Group Evaluation

making a choice as well.' The second lecture by a SN statistician contrasted this assertion. When the presenter was asked whether their visualisations had an explicit political viewpoint, they responded that they left the politics to the public, 'so you [analysts and statisticians] don't have to make choices.' Much like in the brainstorm workshop, what the first lecture introduced was an orientation to social problems, and in this case, the appreciation of an aesthetic to represent and bring attention to these problems. Furthermore, the NGO visualisations were presented as aesthetically more pleasing than those of the NSI. The NGOs were detailed, interactive, applied subtle colour schemes, and were easy to grasp because they were based on geographic maps. The NSI visualisations, although innovative, were clunky, less concerned with continuous and

cohesive colour schemes and, while understandable to statisticians, less easy for publics to read. The data camp mentors encouraged attractive visualisations, as one NSI mentor stated: 'It would be great if we had something like the [NGO] visualisations on our website.' The teams were also coached actively to produce such visuals.

But visualisations were not only encouraged because they could draw in publics; they were also discussed and used as analytical techniques for interpreting large volumes of data that are not easily analysed using traditional techniques such as graphs. That is, the aesthetics of visualisations not only make them 'attractive' but through their use of contrasts, colours, and animations also facilitate analysis. So, while maps, graphs, and diagrams have always been part of statistical analyses, the difference here is the appreciation of the analytic possibilities of advanced aesthetics. Much like algorithms, they help to demonstrate the potential of big data.

A final sensibility was introduced by the NSI mentors in the context of preparing for the closing presentations: experiments and other risk-prone formats as instruments for developing business cases to support innovation projects. Participants were encouraged from the start to not only think in terms of results-oriented projects for specified groups of users, but also to be inquisitive and to take risks. This valuation of the importance of experimentation was underscored by the NSI's Director General in their presentation at the end of the final day of the data camp. As the data camp demonstrated, this included learning to manage the tensions in doing trial-and-error work that may not always lead to the desired results or quality in a short time span. Statements such as 'there is a lot in the data' helped resolve such tensions, as well as suggesting the future potential of a project.

The brainstorming workshop and the data camp are professionalising practices that made explicit dispositions of

embodied forms of cultural capital: a feel for the business case and users; the aesthetics of visualisations; experimentation; patience; and an affinity for and appreciation of algorithms. Rather than all-encompassing or constituting a universal data science habitus, these are some of the acknowledged sensibilities that make up an entrepreneurial habitus required to recognise the future potential of big data. While embodied by data scientists working in the technology sector, the sensibilities are valued for their capacity to solve social problems through the uptake of big data such as that generated by social media. It is through such valuations that it can be said that big data and the entrepreneurial habitus of data scientists are at once in tension and compatible with a public service habitus, which entails a commitment to working for the common good. In the following section we explore how this tension plays out in the professionalising practice of conferences where the entrepreneurial skills and habitus of data scientists were both valued and opposed by differentiating and defending them from those of public service.

Conferences: Defending by Differentiating

As in other fields, the profession of national statistician is shaped and defined through complex interactions and exchanges, from small meetings and official documents to those of international task forces and conferences. Amidst calls to embrace novel working practices, such as the data camp discussed above, statisticians also regularly convene at international meetings to discuss changes and challenges facing their profession. In this section we focus on Eurostat and UNECE conferences where big data and innovation were part of the agenda. The meetings followed a very traditional bureaucratic format of presentations and discussions typically centred on

PowerPoint presentations from authoritative figures in the field. Mundane arguments about the novelty of big data, data science, and so on were often repeated in an uncritical manner. Such repetitions highlight that the skills and habitus of a data scientist emerge not only through material-semiotic practices, but also through discursive ones. In what follows, we examine conference statements and debates as professionalising practices that involve differentiating the skills and habitus of national statisticians not only in line, but also opposition to that of data scientists.

As elaborated previously, data scientists are being defined not only in relation to particular sets of technical and analytical skills (or cultural capital) needed to manipulate large data sets, but also particular embodied sensibilities. For example, at a 2016 UNECE conference, a statistician criticised a lengthy presentation about the impact of big data on official statistics by pointing out that the change needed from NSIs goes much beyond the acquisition of new skills and toolsets:

> The [previous] presentation was very much tool oriented. We are very familiar with all these tools and the thing that was missing from the presentation was an acknowledgment of the fact that what is actually changing at the moment is the paradigm around how we conduct research. With big data you have the data first and then you ask the questions. The issue is therefore not what tools to use but what questions to ask. That's the crux of the matter, and that is where the skills come in.

NSIs are no longer 'the farmers', the presenter continued, but 'foragers of data'. As such, the key concern was 'what questions to pose and how to draw inference' and 'how to produce the best possible estimates to meet user needs from multiple data sources'. For other speakers, it was urgent that statistical

agencies shift their focus from producing statistics to a 'more service-oriented attitude … to connect, aggregate and tailor' statistical information based on user needs and to do so increasingly. 'Service orientedness' was an often-repeated term, which is defined in a number of ways. At this particular conference a consensus seemed to exist that 'service orientedness' refers to 'value added' activities such as analysing and interpreting data, rather than a narrow conception of the NSI role as data collector (UNECE, 2015: 4). In sum, big data is seen to disrupt not just established methods and techniques, but an entire paradigm of producing statistics, which also requires new sensibilities – for example, what questions to ask – and new skills – for example, what valued added activities to deploy.

For some statisticians, the appropriate response to what they conceive of as the challenge of big data is that NSIs need to become more like their private sector competitors. For example, at a meeting organised by Eurostat in 2016, a senior manager explained that not only do private companies now accumulate vast amounts of big data, they have the 'mindset of a big data company':

> The big advantage they [Facebook and Google] have is that they have the big data to accomplish a maximum effect. They also have the mindset of a big data company, which the statistical community does not. When we started using administrative data at [our NSI] statisticians were violently opposed to them with fundamental principle reasons. The same thing is happening with big data.' 'This is not statistics, this is not quality', they say. The first thing to do, therefore, is to get the mindset right.

The move from a product to service orientation was identified as involving a cultural change at NSIs, one that must begin at the very top level of managers. At a practical level, the shift in

mindset referred to in the above quote was conceived of as involving a willingness to accept different definitions of quality, since the sources from which data are derived are becoming increasingly varied. That NSIs look to the private sector for examples of adopting a more service-oriented approach is perhaps unsurprising. Amidst increasing competition, the 'modernisation' of statistics often refers to the adoption of a private sector mindset or entrepreneurial habitus.

However, the appropriation of an entrepreneurial habitus was not the only response of statisticians to the challenges of big data. Nearly as regularly, the future of the profession was also defined in contrast to values held in the private sector by reinforcing and defending long-held public service values in the production of official statistics. Indeed, while big data raised questions about the skills and competencies of national statisticians, existing values that they command such as trustworthiness, public accountability, civil service and democratic legitimacy were also defended. As in the case of job interviews, conference presentations stressed a public service habitus that values ethics and quality standards involved in the everyday production of official statistics. Such valuations occur, for instance, when some statisticians ethically objected to the use of corporate data sources because they cannot verify their quality according to formal standards (Struijs, Braaksma, and Daas, 2014).

These values constitute another repetition often asserted at international conferences: that the investments of NSIs in myriad forms of data and their capacities to secure the principles of official statistics ensure the relative advantage of national statisticians in the future. As stated in a paper presented at a UNECE conference in 2013, official statistics have a 'trademark' based on quality criteria that need to be protected:

> It is unlikely that NSOs [NSIs] will lose the 'official statistics' trademark but they could slowly lose their reputation and relevance unless

> they get on board. One big advantage that NSOs have is the existence of infrastructures to address the accuracy, consistency and interpretability of the statistics produced. By incorporating relevant big data sources into their official statistics process NSOs are best positioned to measure their accuracy, ensure the consistency of the whole systems of official statistics and providing interpretation while constantly working on relevance and timeliness. The role and importance of official statistics will thus be protected (UNECE, 2013: 2).

Statisticians, in other words, asserted their authority to establish, but also to evaluate adherence to, quality criteria in the production of official statistics. Thus, while the effects of big data are considered disruptive, it affords the opportunity to defend the relative advantages of the official statistics and the skills and habitus of national statisticians. Data scientists were not 'taking over' or replacing statisticians but were differentiated from national statisticians. In other words, while requiring new skills, big data is also (potentially) reinforcing established values and norms.

Yet again, like the different positions taken on the challenges of big data, counter arguments were also advanced about the extent to which NSIs can hold on to such traditional values in the midst of increasing competition between data producers. At a 2016 UNECE conference this came up in relation to discussions of data ethics. Responding to a presentation about the numerous potential ethical issues concerning NSIs using big data, a statistician made the point that even total abstinence would not free NSIs from ethical concerns. For them, this would only result in big data being left solely in the hands of actors who care less about ethical considerations than statisticians:

> I am concerned about finding the right balance. In your assignment, you have explored all potential objections to using big data in official statistics. But there is also an ethical concern with us not engaging with the data, because even if we did not use them, others still would.

> For example, we have been experimenting using Twitter data, and our legal experts have been complaining to us about it. But individual social data is already on the market. Individual psychological profiles can be purchased from social media companies. This is the reality, and in this reality we cannot be too strict about ethics.

In other words, increasing competition from different private sector data producers raised a concern whether NSIs can hold on to their long-held principles such as those related to data ethics in the context of a 'new reality'.

This, as in the other professionalising practices, makes explicit tensions between entrepreneurial and public service skills and habitus. Whether discursive or material-semiotic, such tensions are manifest in multiple ways and how they play out and their consequences for the statistician subject and profession of national statistician are by no means settled or certain. Rather, they are objects of struggle over recognised forms of cultural capital (skills) and habitus (embodied dispositions) that are valued and recognised in the transnational field of statistics. It is to that point that we turn in the conclusion.

Conclusion

This chapter shifted attention to professionalising practices to understand how data practices not only involve struggles over methods of producing statistics. To advocate particular data practices also involves valuing the skills and habitus required to perform them and in turn the relative advantages that may be conferred to professionals who possess them. We highlighted how such valuing happens through both discursive and material professionalising practices. Acknowledging that there are numerous professionalising practices (e.g., training programmes, university curricula etc.), our aim is to exemplify

one aspect: how they involve a tension between the cultivation of entrepreneurial and public service skills and habitus.

This tension was evident in the professionalising practices that we analysed. Whereas the workshop and data camp blurred the boundaries between addressing social problems through technological innovation (an entrepreneurial positioning) and working for the common good (a public sector positioning), the conferences engaged in boundary making around values. With regard to the role of official statistics in the changing landscape of data production, this is likely to be an enduring tension. For example, business and political concerns were publicly raised about SN's use of private sector data, its increasing presence as a market competitor for work commissioned by businesses, and its uptake of predictive methods. In response, new regulations were adopted in 2020 that stipulated that SN would primarily produce statistics for the public sector (Brasser, 2019; Minister van Economische Zaken en Klimaat, 2020). What this exemplifies is that entrepreneurialism, innovation, and values of public service will likely continue to be objects of struggle within a political economy of data production.

However, following the statement from SN's Acting Director General that 'we'll still contribute to major social issues such as energy transition, sustainability, poverty and debt problems', among others through 'data-driven working and innovation', entrepreneurial and public service dispositions remain closely aligned for this NSI (Statistics Netherlands, 2020). Moreover, it demonstrates the relevance of professionalising practices: they play a role in cultivating the priorities and values relevant for how NSIs are positioning themselves in a changing landscape of data producers and shape the practices they adopt to produce official population statistics.

References

Bigo D (2011) Pierre Bourdieu and International Relations: Power of Practices, Practices of Power. *International Political Sociology* 5(3): 225–258.

Bigo D (2014) The (In)Securitization Practices of the Three Universes of EU Border Control: Military/Navy – Border Guards/Police – Database Analysts. *Security Dialogue* 45(3): 209–225.

Bourdieu P (1984) *Distinction: A Social Critique of the Judgement of Taste*. Cambridge, MA: Harvard University Press.

Bourdieu P (1989) Social Space and Symbolic Power. *Sociological Theory* 7: 14–25.

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.

Brasser P (2019) Hoe commercieel mag het CBS opereren? *Het Financieel Dagblad*. 13 December.

Gehl RW (2015) Sharing, Knowledge Management and Big Data: A Partial Genealogy of the Data Scientist. *European Journal of Cultural Studies* 18(4–5): 413–428.

Grommé F, Ruppert E, and Cakici B (2018) Data Scientists: A New Faction of the Transnational Field of Statistics. In: Knox H and Nafus D (eds) *Ethnography for a Data Saturated World*. Manchester: Manchester University Press, pp. 33–61.

Halavais A (2015) Bigger Sociological Imaginations: Framing Big Social Data Theory and Methods. *Information, Communication & Society* 18(5): 583–594.

Halford S and Savage M (2010) Reconceptualizing Digital Social Inequality. *Information, Communication and Society* 13(7): 937–955.

Haraway DJ (1997) *Modest Witness@Second Millennium. FemaleMan© Meets OncoMouse™*. Feminism and Technoscience. New York and London: Routledge.

Irani L (2019) *Chasing Innovation: Making Entrepreneurial Citizens in Modern India*. Princeton, NJ: Princeton University Press.

Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.

Lahire B (2012) *Monde Pluriel: Penser l'unité Des Sciences Sociales*. Paris: Seuil.

Latour B (1993) *The Pasteurization of France*. Cambridge, MA: Harvard University Press.

Latour B (2004) Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern. *Critical Inquiry* 30(2): 225–248.

Mackenzie A (2013) Programming Subjects in the Regime of Anticipation: Software Studies and Subjectivity. *Subjectivity* 6(4): 391–405.

Meulman J (2016) When It Comes to Data, Size Isn't Everything. *STAtOR* 17(2): 37–38.

Minister van Economische Zaken en Klimaat (2020) Beleidsregel taakuitoefening CBS. WJZ/19207028, Den Haag: Ministerie van Economische Zaken en Klimaat.

Muniesa F and Linhardt D (2011) Trials of Explicitness in the Implementation of Public Management Reform. *Critical Perspectives on Accounting* 22(6): 550–566.

Naur P (1974) *Concise Survey of Computer Methods*. Sweden: Petrocelli Books.

Pasquale F (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Ruppert E and Scheel S (2019) The Politics of Method: Taming the New, Making Data Official. *International Political Sociology* 13(3): 233–252.

Ruppert E, Law J and Savage M (2013) Reassembling Social Science Methods: The Challenge of Digital Devices. *Theory, Culture & Society*. Special Issue on 'The Social Life of Methods.' 30(4): 22–46.

Savage M (2010) *Identities and Social Change in Britain since 1940: The Politics of Method*. Oxford: Oxford University Press.

Scheel S, Grommé F, Ruppert E, et al. (2016) *Transcending Methodological Nationalism through Transversal Methods? On the Stakes and Challenges of Collaboration*. ARITHMUS Working Paper 1. Available at www.arithmus.eu (accessed 1 April 2020).

Shapin S and Schaffer S (1985) *Leviathan and the Air Pump: Hobbes, Boyle and Experimental Life*. Princeton, NJ: Princeton University Press.

Statistics Netherlands (2020) New regulations direct the role of CBS. Statistics Netherlands Corporate. Avail. At www.cbs.nl/en-gb/corporate/2020/27/new-regulations-direct-the-role-of-cbs (accessed 17 April 2021).

Steinmetz G (2016) Social Fields, Subfields and Social Spaces at the Scale of Empires: Explaining the Colonial State and Colonial Sociology. *The Sociological Review* 64(2). SAGE Publications Ltd: 98–123.

Struijs P, Braaksma B and Daas PJ (2014) Official statistics and Big Data. *Big Data & Society* 1(1): 1–6.

UNECE (2013) *What Does 'Big Data' Mean for Official Statistics?* 18 March. Geneva: United Nations.

UNECE (2015) *Innovative Products and Services for European Statistics: Challenges and Opportunities*. ECE/CES/2015/42. Geneva: Conference of European Statisticians.

UNECE (2018) *Recommendations for Promoting, Measuring and Communicating the Value of Official Statistics*. New York and Geneva: United Nations.