SPSS as an 'inscription device': from causality to description?

Emma Uprichard, Roger Burrows and David Byrne

Abstract

This paper examines the development of SPSS from 1968 to 2008, and the manner in which it has been used in teaching and research in British Sociology. We do this in order to reveal some of the changes that have taken place in statistical reasoning as an inscription device in the discipline over this period. We conclude that to characterise these changes as a shift from 'causal' to more 'descriptive' modes of analysis is too simplistic. Such a shift is certainly apparent, but it meshes in complex ways with a range of other – just as important – changes, that together mark a phase-shift in the functioning of sociological quantification.

Introduction

From the late 1960s onwards, British sociologists have had access to a large number of different statistical software packages. It is difficult to estimate the exact number that have been used, but there have probably been 50 or more systems that have been utilised for teaching and research at different times over the past 40 years or so. Some, such as *GLIM* (Generalized Linear Interactive Modelling) developed by the Royal Statistical Society's Working Party on Statistical Computing, were popular for a time, but then fell out of favour. *Minitab*, was popular for teaching in the 1980s and remains so in some institutions. In addition, a recent 'rough and ready' audit of quantitatively inclined colleagues generated the following list of packages used in a sustained manner in recent decades: *LISREL*; *MLwiN*; *R*; *SAS*; and *Stata*; some colleagues also noted how widely *Excel* is now used for basic quantitative analysis. Readers, undoubtedly, will be able to add to this list. There has been one package, however, that has not only endured but has also remained the most widely known and used: the seemingly ubiquitous *SPSS – the Statistical Package for the Social Sciences*. So central to the experience of 'becoming' and 'being' a sociologist in Britain over the last few decades, neither the material and the semiotic functioning of this particular piece of software nor the paraphernalia
.

surrounding it (manuals, textbooks, courses and the like), can be ignored if we are fully to understand the panoply of inscription devices that constitute sociological forms of knowledge.

For those familiar with using SPSS in their sociological teaching and research over a number of years, but who have not paid much attention to the changing context of its development and dissemination, a visit to www.spss.com in 2008 might come as something of a surprise.[1] First, the corporate history section[2] makes clear that, although SPSS 's*tood for* the Statistical Package for the Social Sciences' (emphasis added), it is now no longer an acronym but a named software *brand*. Indeed, one has to drill down deep into the site to recover this etymology. Second, for those who might, at one time or another, have felt some sort of vague sociological 'responsibility' for the software, this tenuous disciplinary connection with one of the core 'tools of our trade' is quickly obliterated when it becomes apparent that the site functions to interpellate not 'social scientists', but those seeking 'business solutions'. One quickly feels naïve to have imagined that in an era of 'knowing capitalism' (Thrift, 2005) some of the core tools of the social sciences would not have been fully 'commercialised' and 'globally branded' in this way. Naïve or not, the profound and stark transformation of SPSS from a tool for empirical social research to a corporate behemoth primarily concerned with something called 'predictive analytics' (about which more below) has been fast and dramatic.

Our story of the change in British quantitative sociology is therefore one that is told alongside changes to SPSS and its impact on ways of knowing the social world. After all, if it is the case – as has recently been argued (Savage and Burrows, 2007) – that we are at the cusp of a crisis of 'causal' forms of empirical analysis in sociology, which results in a need to recover more 'descriptive' forms of the discipline, then we might suspect that this will manifest itself emblematically within the algorithms, interfaces, visualisation tools and other forms of inscription device that SPSS offers up. Indeed, we argue that SPSS Inc. and the numerous products that they now produce under the auspices of the SPSS brand represent not just an early *instantiation* of this shift in sociological orientation, but rather a prefigurative catalyst in bringing it about.

Underpinning both the changes in quantitative sociology and SPSS, however, are the processes of digitization and associated changes in the availability of data. Indeed, we consider *this* aspect of the wider social world to be an essential driver of the various transformations described here. Although the processes of digitization remain a largely implicit strand of our story of change, they are fundamental as they have been quietly making their presence felt in quantitative research more generally, so they cannot be ignored here either.

This overall argument emerges from the research conducted to prepare this paper. At the outset, we worked out a crude schema for periodising the recent history of quantitative methods in British sociology drawing on numerous historical sources, ranging from general accounts of change, such as Abrams

607

(1968), Hindess (1973), Irvine *et al.* (1979), Kent (1981), Bulmer (1985) and so on, right through to some of the Institute of Sociology's reports on teaching across the social sciences. Because it became clear that the effects of the PC were important to the changes visible in sociology, we then 'mapped' this onto developments in both hardware and software for statistical analysis. This took us to the corporate history section of the SPSS Inc. website, which offers a reflexive account of how SPSS Inc. came to possess a wide portfolio of software, functions and customers. The website offers both a narrative history and a very interesting attempt to periodise the institutional history of the brand. Ironically, however, our own periodisation of sociology and developments in statistical analysis and the narrative account offered up by SPSS Inc. proved to be almost exactly homologous! In what follows, therefore, we describe SPSS Inc.'s self-periodisation and how it meshes with the post-1968 history of British quantitative sociology.

SPSS is born

Before SPSS and some other packages became available, researchers had no choice but learn to use the high level programming language Fortran to write their own programmes and relied on the likes of Veldman's (1967) *Fortran Programming for the Behavioral Sciences* to help them do so. SPSS was developed in 1968 by Norman Nie, Tex Hull and Dale Bent – all then at Stanford University – as an alternative Fortran based program, specifically designed to analyse quickly the large amounts of quantitative social data being gathered by Faculty and Graduate students. Initially, therefore, SPSS was primarily aimed at allowing researchers to do basic descriptive statistics, cross tabulations and regression. Other packages were used for other purposes; for example cluster analysis tended to be carried out using CLUSTAN. However, once news that SPSS was available for basic statistical analysis, what was colloquially known as 'the Stanford Package' was soon in demand in other US institutions.

In 1969, Nie and Hull moved to the University of Chicago – Nie to the National Opinion Research Center and Hull to the University's Computation Center. A year later, the publishers McGraw-Hill repackaged the documentation that had been produced to accompany the software as the first SPSS User's Manual. This sparked a huge demand for the programme and the income generated from the royalties from sales of the manual was substantial enough to threaten the non-profit status of the University of Chicago; so in 1975, SPSS became incorporated as a business with the two founders, Nie and Hull, as the company's executives.

The first versions of SPSS available in the UK were written in Fortran, so users still required quite high level programming skills. They ran on mainframe computer systems stored in large air conditioned rooms, often large enough to be small lecture theatres. The mainstreaming of such computerization made it

608

possible for quantitative social researchers to radically speed up analytical procedures. Nevertheless, throughout the 1970s most researchers in Britain still used variations of the Hollerith Punch Card system (Grier, 2005). This was one of the first mass data storage and sorting systems, and it employed index cards marked with holes, which acted as a form of code. Later counter-sorters became available which facilitated the use of cards that could be multi-punched. Sara Arber, currently a Professor of Sociology at the University of Surrey, and one of the sociological pioneers of secondary analysis (Dale *et al.*, 1988) was a very early user of the system, and in a personal communication to the second named author, she relates some of her recollections. Importantly, they are remarkably resonant of a number of others we heard whilst preparing this paper:[3]

> I first used SPSS in the US, when I was a graduate student in 1973–74 at Michigan, using punch cards (of course). I got a one year Lectureship at Surrey starting in September 1974 to teach methodology, and one of the things that Asher Tropp [the Founding Head of Sociology at Surrey and a key figure in the development of sociological research methods in the UK] was particularly keen on was teaching students to analyse 'real' survey data using SPSS. So, in 1974–75, I started teaching undergraduate students how to use SPSS and we used the US General Social Survey . . . At this time, all the Surrey computing was via an over night link to the computer in Guilford Street, University of London. Students used to punch their cards for an SPSS run (in a dedicated punch room in the computer centre at Surrey) and then the cards were submitted to London – the following day you would receive the paper output (either showing an error or some analyses, if you were lucky). So, it was very frustrating for students and everyone else to have to wait overnight for any results . . . At this time, I was unaware of any UK survey datasets that could be used for analysis using SPSS.

These personal recollections highlight a number of significant issues that are important when trying to understand the history of British quantitative sociology, which are worth spelling out more explicitly.

First, it is worth noting the significant role that the Department of Sociology at the University of Surrey played in the direction quantitative sociology has taken. Its early MSc in Social Research, along with its short training courses, placed Surrey as one of the first professional training grounds in sociological research. Surrey, alongside the University of Essex Summer Schools in Social Science Data Analysis & Collection, was *the* major conduit for the promulgation of developments in statistical analysis in the discipline. Surrey was also crucial in making available large-scale official data sets in easy to use SPSS versions. The work that Sara Arber, Nigel Gilbert and Angela Dale did on converting General Household Survey (GHS) data files into fully documented SPSS systems files (Gilbert *et al.*, 1983),[4] was catalytic in transforming both teaching and research in the UK.

609

Second, the process of *doing* quantitative research has changed significantly. Both data and programming instructions had to be entered on cards which were punched by operators from printed sheets which researchers had to fill in painstakingly for themselves. One card was one row of data; one mistake on the card meant throwing it away and re-punching the entire card. Data often had to be copied from hundreds of questionnaires or from printed sources. This was all simply to produce the data set, as there were then, of course, no online resources, although sometimes data could be obtained as card sets or, later on, as magnetic data tapes (the 1971 UK Census was amongst the first data sets to be available in this format) that could be mounted on large tape readers.

As Sara Arber noted, the mode of command and data entry and the relatively slow speed of computing at that time meant that turn around was at least overnight or longer; more frequently, it would often take much longer than anticipated, since waiting for a line of 'errors' was part and parcel of doing statistical analysis. The third named author remembers having several abortive runs until he had worked out that he had used exclamation marks instead of the number one. Mistakes such as these were easy to make given the necessary precision involved in entering data. For example, cases which might go onto multiple lines of a card if there were many variables had to be specified in a manner that now appears arcane: 10(F8.0)/4(F8.0,F6.2,F5.2,F1.0) . . . Output came on printed sheets from line printers – hundreds of them.[5]

Although the Hollerith punch card system and counter-sorters transformed quantitative collection and analysis, compared to today's computing standards, it was incredibly slow. It also demanded a reasonable level of statistical understanding, along with practical skills in data entry and command language. As obvious as it may seem, it also required the ability to *type*. Although by the 1970s typing was pretty well a universal skill among Americans, British academics still often wrote with pen and paper and had secretaries type up their work, so in fact many academics, especially male ones, could not type. Indeed, if one can read beyond the acerbic critique of the discipline contained in Malcolm Bradbury's (1975) *The History Man*, much of the description of work-a-day life contained within it does include a fairly accurate portrayal of the technologies and divisions of labour that pertained in British sociology in the early 1970s. All in all, doing *any* kind of quantitative work was incredibly labour intensive. This had implications on *who* had the resources to conduct large scale surveys, which in turn placed government as the sole large-scale survey provider.

Finally, as echoed in Arber's recollections, by the late 1970s and early 1980s, although SPSS was still for UK academics primarily a denizen of their own or other Universities mainframe computers, access to these computers was often done remotely, largely through UNIX command language initially via dumb-terminals, but later from desktop PCs able to act as terminal emulators. Now command file sets could be written electronically and stored for correction. Similarly, large electronic data sets, such as the Surrey GHS files, slowly

610

became available. Operations were faster, even though they were still initially based upon batch processing; commands and data would be entered but were then often queued and submitted overnight. Later still, mainframe versions of SPSS (and later SPSSX) became interactive; one could submit commands and data and, after a time, output could be delivered back to the screen. At the time this was a revelation; it represented a step-change in the speed at which analyses could be carried out.

The 'PC turn'

It was only in the mid-1980s that SPSS introduced the first mainframe statistical package able to run on a PC (versions were shipped able to run on IBM PCs, PS/2 running OS/2, Apple Macintosh, UNIX and VAX/VMS workstations, minicomputers and larger systems, as well as mainframes). SPSS/PC+ ran under Microsoft DOS but still required data and command files to be carefully entered using, what now appears to be, esoteric SPSS syntax but which, at the time, became a central inscription device familiar to all those undertaking quantitative social research. But the shift to the PC was incredibly important, and during the period between around 1984 and 1992, SPSS further consolidated its position as the 'industry standard'.

This period also marks the beginnings of an important *qualitative* shift – a phase shift even – that occurred in the development, pedagogy and use of quantitative methods in British sociology. This is when the computerisation of statistical teaching in Higher Education really took off and the teaching space of the 'PC lab' began to emerge. Note, however, that the pedagogical shift involved more significant, albeit subtle, changes in emphasis on how to interpret, analyse and understand quantitative data. As we imply throughout the remainder of the paper, this shift not only resulted in a different way of *doing* quantitative research, but in a different *kind* of quantitative research as well. In effect, prior to the arrival of SPSS on the PC, there had been a lot more concentration on both the philosophical nature of data – qualitative and quantitative – and with it, the role of the researcher in interpreting and constructing quantitative data. In contrast, post-1980s quantitative pedagogy places more time and focus on the output than the labours of the researcher.

To some extent, a shift in pedagogy was to be expected – students and their teachers needed to learn how to conduct statistics using a computer and related software. In turn, there was a rapid growth in the production of SPSS (and, to a lesser extent, Minitab) guides and workbooks on 'how to *do*' quantitative analysis using the software packages. But what accompanied this shift in *doing* quantitative analysis with a particular software package was a perhaps surprising *substantive* shift in *how* to do it. Textbooks and student guides shifted to having much less discussion on the processes and theories involved in *interpreting and constructing* quantitative representations of the social world than they had done before the widespread use of the software.

Even when the emphasis was not on, say, SPSS *per se*, the focus was still much more about 'techniques' than it was about the 'construction' of quantitative social data.

The shift in how quantitative methods were taught is an important one. The 'black box' nature of SPSS quietly transformed the user as they became increasingly dependent on the software and computer technology to think statistically and to perform analysis. Students were, as a matter of course, taken out of the dialogic space of the classroom and into the more individualised, task performing spaces of the computer lab. Here, they learnt which buttons to click, which menus to use and which parts of the output to focus on. The absolute physicality of the task involving, as it did, working with cards, machines, tapes and huge amounts of paper as *material* instantiations of cases, variables and so on, became absent and was replaced instead with speed and the apparent *dematerialization* of both data and the mechanics of analysis. This became even more the case post-1992 as SPSS migrated to Microsoft Windows and the use of SPSSX on mainframes dwindled.

This tendency to 'do quantitative research *with* SPSS' has been buttressed by the publication of a plethora of textbooks and guidebooks on how to conduct traditional statistical analyses using SPSS, where the focus is first and foremost on the basics of navigating the student through SPSS windows, data inputting and editing, through to how to conduct various statistical procedures and how to interpret them. The task of 'managing the software' has largely been at the expense of discussing more philosophical concerns and the decision making processes involved in making substantive interpretations about the analytical findings.

This is not to say that contemporary statistical textbooks are devoid of theory; many are of course still theoretically driven. However, it is possible to think of quantitative sociology books appearing after the 'PC turn', as falling into one of three camps: those that focus on understanding and interpreting quantitative data at the expense of almost any explicit SPSS guidance, even though it might be referred to in the text (eg Marsh, 1988; Byrne, 2002);[6] then [1] there are those at the other extreme, that are almost entirely SPSS driven to the point that they end up being first and foremost guidebooks on how to use SPSS *and then* accounts of how to conduct quantitative research (eg Field, 2007; Pallant, 2007); and finally, those that lie somewhere in the middle, attempting to combine the interpretation of quantitative data with some brief guidance on what to click and what parts of the output to focus on (Fielding and Gilbert, 2006; Bryman and Cramer, 2008).

Interestingly, while in sociology teaching and learning broadened themselves out by developing these three alternative pedagogic pathways, we see something similar in the way that SPSS developed as well. Up until about 1994, SPSS had remained focused on the production of its own statistical products. The Windows version of SPSS, for example, generated huge income for SPSS Inc. – reaching some $84m by 1996 according to the SPSS website. However, assisted by the commonality of the Windows interface, it began to acquire

other statistical products from other companies which it attempted to integrate into the overarching SPSS brand. From a commercial point of view, this was a success: again according to their website, by 1997 SPSS had revenues of $110m. Clearly, this was no longer an academic small business; it had become a global corporation.

Towards 'predictive analytics'

The process of acquisition accelerated post-1997 as SPSS took over a number of different software companies. Instead of acquiring products similar to the ones that had long been at the core of SPSS as a piece of 'social science' software, the products that were now being sought were ones that offered 'data mining', 'business intelligence', 'web analytics', 'online analytical processing' and 'text mining'.

On the one hand, then, SPSS is still concerned with some things most sociologists will recognise: market research, survey research, public health, administration and institutional research, statistics, structural equation modelling, education, government and health. On the other hand, there are other items, many foregrounded over those just mentioned, with which most sociologists may be less familiar: marketing effectiveness, fraud detection and prevention, risk management, enterprise feedback management, business intelligence, data mining, text mining, web analytics, financial services, insurance, retail, telecommunications and, at the top of the list, predictive analytics. According to SPSS, this shift in software development was to meet an 'expanding need for understanding ever-increasing volumes of data, and to support- . . . the widespread use of data in decision-making.' By 2002, revenues had almost doubled to $209m.

Our sense, as routine users of SPSS, is that within British sociology this change in direction has not really registered. Most social researchers have continued to use SPSS in the manner to which they have become accustomed, primarily for the analysis of primary survey data and the secondary analysis of large data sets using a fairly routine set of statistical procedures. Some approaches have gained in popularity with improvements in computational power – logistic and probit regression analyses for example (reliant as they are on iterative algorithms) – and some methodological advances – multi-level models for example – have become mainstreamed. Few in the British sociological heartlands, however, appear to have taken cognisance of the implications for the discipline of the profound processes of social digitization that are occurring (McCue, 2007; Savage and Burrows, 2007; Thrift, 2005). Yet this shift has been fundamental to driving the strategic shift that SPSS Inc. have made towards, what they articulate as, the 'age of predictive analytics'. As McCue neatly sums up:

> Whether it is called data mining, predictive analytics, sense making, or knowledge discovery, the rapid development and increased availability of

advanced computational techniques have changed the world in many ways. There are few, if any, electronic transactions that are not monitored, collected, aggregated, analyzed and modelled. Data are collected about everything from our financial activities to our shopping habits. (McCue, 2007: xxv)

Not surprisingly perhaps, SPSS Inc. has quickly caught on to the implications that the increased digitization of data implies, and it has successfully established a market segment in predictive analytics. Indicative of this shift was their acquisition of the Netherlands-based DataDistilleries towards the end of 2003.

So far, the turn towards predictive analytics in quantitative research has taken hold primarily within the business and policy sectors, whilst the impact on British sociology has been muted.[7] On the one hand, as a *substantive topic*, it has been a central feature of what has come to be known as 'surveillance studies'. Sociologists have voiced concerns about the impact of the implementation of such technologies conceptualised as 'software sorting' devices able to produce a 'phenetic fix' on society (Graham, 2004; 2005; Lyon, 2002; 2003; Phillips and Curry, 2002). On the other hand, however, the *methodological implications* of these technologies for sociological practice have amounted to little more than vague rumblings (Savage and Burrows, 2007) about the possibilities that they afford. Of course, these rumblings have not been articulated in terms of 'predictive analytics' – this would smack of a commercial sensibility many within British sociology would find objectionable – but the methodological *logic* behind the approach has certainly been explicitly referred to. For example, some have suggested that geodemographic and socio-spatial profiling offer some new substantive and methodological opportunities (Burrows and Gane, 2006; Parker *et al.*, 2007; Savage and Burrows, 2007) in relation, for example, to debates about the spatialisation of social class (Savage *et al.*, 2005). Others have pointed towards the importance of describing and exploring data to identify particular *types of cases* – a key element of predictive analytics. For example, Chapter Six of Byrne's (2002) *Understanding Quantitative Data* is precisely about 'exploring, describing and classifying' which, it is argued, are key to studying the complex social world.

Although 'predictive analytics' might seem like an uneventful turn of events, we suggest that they are emblematic of a new methodological ethos in quantitative methods. Indeed, we argue that whereas the mid-to-late 1980s marked a first phase-shift relating to the increased use and availability of the PC and related software, today the course of British quantitative social research is witnessing a second radical shift. This is certainly being driven by the increasing ubiquity of digitization processes but, relatedly, also by changes in domain assumptions about contemporary social ontology. For the business sector, this new methodological ethos is driven, primarily, by the cold logic of the profit motive; in sociology, it comes via the recent 'complexity turn' (Byrne, 1998; Urry, 2003), which involves a quite fundamental reappraisal of a number

of domain assumptions that, potentially, involve nothing less than a 'paradigm shift'.

In what follows, we offer an *over-characterisation* of this methodological shift and what is at stake here. We do this for two reasons. First, to keep engaged those few readers with no inherent interest in SPSS, but an active interest in some of the conceptual aspects of the paper hitherto. Second, as a methodological device, to present a stark ideal-typical contrast to which we hope colleagues will react, in a further attempt (Savage and Burrows, 2007; Byrne, 2002) to jolt the discipline out of a methodological complacency that the 'coming crisis' will exploit mercilessly to the detriment of us all, unless we confront, in a pro-active manner, the implications of the new realities of 'knowing capitalism' (Savage and Burrows, 2007; Thrift, 2005).

The 'new face' of quantitative research?

Most sociologists in Britain have not explicitly connected what we are going to characterise as the emergent 'new face' of quantitative research with the rise of the digitization of routinely constructed transactional and administrative data (Savage and Burrows, 2007) or the 'everyday life' data banks that 'Web 2.0' technologies supply (Beer and Burrows, 2007). Yet this 'digital turn' in the availability of data is precisely what makes the SPSS rhetoric of the 'age of predictive analytics' both possible, and some of the methodological shifts associated with it, potentially obligatory. We summarise some of what follows in highly schematic form in Table 1.

The emergent, or 'new face', of British quantitative sociology operates at two levels. The first relates to the actual techniques that have been 'in vogue' at various points over the post-war period. This is a relatively easy task to narrate and, indeed, we have already made a start on this in the account of SPSS we have already provided. Over time some statistical techniques have become favoured over others. This is not surprising – things change, fashions come and go, new methods emerge and new technologies offer new affordances. However, SPSS and other quantitative analysis software packages are prime examples of a set of inscription devices that notoriously set some techniques as 'default' options, which may or may not easily be altered to allow others to be foregrounded. Let us take, by way of a simple example, the stem-and-leaf plot.[8] This is one of Tukey's (1977) many graphs used to describe a single continuous variable that is, in many respects, more informative than its histogram cousin. Like the histogram, it displays the overall shape of the distribution, but it also provides a very precise display of the values that give the distribution its shape. The stem-and-leaf plot is available in SPSS, but unlike the histogram, it is not readily available via the upfront 'graph' menu. Instead, it is hidden away as a tick box under 'options' in 'descriptive statistics'. What is set as 'default' in SPSS, and indeed the way it is provided to the user, then, affects the interpretation and use of each and every technique. The stem-and-leaf plot – as is

Table 1 A schematic overview of changes in British
quantitative sociological research

Shift	'Old Face'	Emerging 'New Face'?
Technique	• Parsimonious model building, eg multiple regression analysis, factor analysis, multi-level analysis, etc.	• Describing and exploring groups, eg cluster analysis, simulation (micro-simulation and agent-based simulation especially), correspondence analysis, QCA, etc.
Generalisation	• Generalisation as possible, reliable and desirable	• Move to 'moderatum generalisation' which are somewhat reliable, but remain more or less enduring relative to future research and findings
Causality	• Single causes • Linear causal models • General causal 'laws' – outcome consistently explained through particular variable interactions • Faith in 'finding' causes	• Multiple, contingent causality • Configurational causality – same outcome possible through different variable configurations; different outcomes possible through same variable outcome • Complex, nonlinear and in flux • Less faith in 'finding' causes
Prediction	• Onus on predicting the ultimate single future • Considered possible	• Onus on describing multiple possible futures instead of determining one simple future • Possibility of prediction becomes questionable
Sampling	• Probability sample as best possible form of knowing population • Sample used to statistically infer sample findings to population • Statistical significance testing as key to understanding population	• Population data widely available through increased digitization of data • Statistical inference is seen as unnecessary • Probability sample used to confirm descriptions of population rather than inferring to them
Interpretation	• Focus on explanation and confirmation	• Focus on description, exploration, classification, case profiling and visualisation.
Variables and Cases	• Focus on the variable	• Focus on the case and describing types of cases.

portrayed through SPSS – is seemingly less useful than those other inscription devices that are readily available via the Graph menu.

Of course SPSS has not evolved in a vacuum. Nor has it broadened its analytical capacities randomly. Inscribed in SPSS itself is a long history of 'analytical memory' which is visible as each new version brings with it its own story of what users might now want and need, and of course what is technologically possible. Some techniques, such as social network analysis, are still absent from SPSS whereas others, such as correspondence analysis and cluster analysis, have become 'mainstreamed'. Conversely, basic new mechanisms become available as SPSS becomes more 'interactive'. Thus, for example, version 12.0 came with (among other things) a new 'visual bander' tool (now called 'visual binning' – the argot itself is not unimportant here, and elsewhere, of course). This allows for a continuous variable to be displayed as a histogram, which can then be interactively 'cut' into groups to produce a more meaningfully data driven ordinal variable.

There are many more possible examples, but the point remains the same: the inscription devices involved in quantitative software developments are an intrinsic part of the way that quantitative sociological knowledge is constructed. As Law (2003) explains, inscription devices facilitate the emergence of specific, more or less routinised, realities and statements about those realities. But this implies that countless other realities are being un-made at the same time – or were never made at all. To talk of 'choices' about which realities to make is too simple, and smacks of a voluntarism that does not, in actuality, pertain. The hinterland of standardised packages at the very least shapes our 'choices', but in addition, what is 'supplied' as a 'choice' by companies like SPSS Inc. depends on the 'needs' and 'demands' of its (increasingly commercial) users.

The second level at which the 'new face' of British quantitative sociology operates relates to the epistemology and ontology of the social world after digitization, which we can do little more than register here and signpost for future discussion. The data inundation that has resulted from the digitization of routine transactional operations offers the empirical basis for ever more strategic, reflexive and intelligent forms of 'knowing capitalism' (Thrift, 2005). Moreover, the 'new face' also symbolises changes in the fundamental 'nuts and bolts' of quantitative sociology; our assumptions about, *inter alia*, 'causality', 'prediction', 'generalisation', and last but not least, 'variables and cases', all require a thorough interrogation in the light of, what are, changed circumstances.

With respect to *causality*, for example, the search for 'causal laws' or at the very least 'causal tendencies', has been a longstanding characteristic of quantitative research. Various models have emerged over time – Cartwright (2007) estimates around thirty – the most common being the Humean models of cause and effect as a temporally ordered constant conjunction. More recently, however, there are signs of an alternative approach to causality, which is considered to be complex, contingent, multidimensional and emergent (Byrne, 1998, 2002). This alternative approach to causality has resulted in quite different modes of 'knowing the social', such as multi-agent based simulation

617

modelling (Gilbert and Troitzsch, 2005; Gilbert, 2008) – which do not (yet?) feature in SPSS. In such approaches, 'causal powers' are *multi*-dimensional, *multi*-directional, and *nonlinear*, and, importantly, they are not necessarily separable from their effects. In other words, the world is much 'messier' than the linear models of mainstream statistical analysis – however nuanced – have so far allowed for.

In turn, the possibility of *prediction* is significantly challenged. In its place is a recognition that history plays a fundamental role in how things change; the present is dynamic and always *becoming*; the future is non-deterministic, but is not random either – the concept of 'multiple possibilities' that are more or less likely is preferred to any linear model of prediction. Knowing the *kinds of trajectories of specific kinds of cases* is considered a key part of 'predicting' the odds of how a 'thing' will change in the future. But in order to know the kinds of things that exist – to go back to Law's point about what *classes of realities* there are – we need to *describe* what is (and maybe also what is not) 'out there'. 'Thick descriptions' – as Geertz (1973) would call them – help us to anticipate (as opposed to predict) the general (but not exact) behaviour of particular kinds of cases.

This has knock-on effects for *generalisation* – that is, the ability to infer observations from a sample at one point in time to the population from which it was taken at another. The concept of generalisation, as a grand overarching principle that allows the inference of observations from one context to another, is arguably defunct. Understanding how cases are 'generally' requires an explicit recognition of how they are 'specifically'. The logic of predictive analytics dismisses the ambitious notion of knowing how things are everywhere all the time, and aims instead to know 'enough' to make, what Payne and Williams (2005: 297) call, 'moderatum generalisations' about particular types of cases. 'Moderatum generalizations' – a concept developed for what they argue takes place in qualitative research – relates to tentative claims that are only relatively enduring; they are not intended to hold good over long periods of time, or across ranges of cultures, and they are also somewhat hypothetical inasmuch as further research – statistical or otherwise – may modify or support them. This means that there must also be an acceptance that quantitative 'moderatum generalisations' need to be frequently 'updated' to 'keep up' with the ways in which cases may become different kinds of cases.

This in fact requires us to reconsider the elemental entities intrinsic to all quantitative research: *variables* and *cases*. After all, the types of data mining technologies that SPSS Inc. now foregrounds are, in fact, involved (implicitly?) in a rather different kind of quantitative programme: the development of *case* based rather than *variable*-centred modelling (see Byrne and Ragin, forthcoming). As we have already noted, in British Sociology we have yet to explore fully data mining technologies; however, we see something remarkably similar in its ethos in relation to Charles Ragin's (1987, 2000) development of Qualitative Comparative Analysis (QCA) which signals a prospective perspective on what might be possible.

618

QCA explores the different configurations (of variables) that describe particular outcomes, which is essentially a form of predictive analytics. These kinds of case based methods present an interesting example in which we see the face of the new quantitative programme really play out. They offer a clear break with linear modelling approaches which have dominated traditional causal modelling and a radical change in quantitative thinking. They also dismiss the idea of universalist claims in relation to causal processes, instead emphasizing that the scoping – the specification of the range of time, space and sets of instances/cases – is an essential component of any knowledge claim.

Whilst sociology's quantitative programme has been dominated by rather traditional approaches – expressed to a considerable extent in Britain by Nuffield College led programmes based on linear modelling – the turn to case based methods opens up a real iterative dialogue between qualitative and quantitative forms of evidence. In a sense, QCA and related methods represent a new kind of inscription device within a potentially new configuration of actors developing a new network of knowledge production.

Final comments

In conclusion, our sense is that newly emergent forms of quantitative sociology involve a move from an explanation of how variables work together towards an *exploration and description of different kinds of cases*, and ultimately tentative predictions based on the depictions of those cases. This requires a focus on exploring and describing empirical data – a radical return to Tukey's (1977) extensive work on exploratory data analysis (EDA), as well as Tufte's (1983, 1991, 1997) various efforts towards the 'visualisation of quantitative data'. Both Tukey and Tufte together offer a different kind of school of thought which seems to have come and gone again between the late-1970s and the late-1980s, but which now seems to be resurgent with the new possibilities that new technologies and digital data inundation offer. Indeed, many contemporary data-mining and predictive analytical techniques look to be little more than Tukey and Tufte on both speed and steroids and much more data – much, much more data!

It is tempting to think that this 'shift' in 'how to do' research is simply a story about how *quantitative* research has changed. But of course, we are also now beginning to see something similar happening in *qualitative* research. Atlas-ti and Nvivo are both competing to take centre-stage across British universities, and debates around the ways in which computer assisted qualitative analysis alter the very nature of *doing* qualitative research itself are increasingly raised (Fielding and Lee, 1998; Richards, 1999; Richards, 2002). Moreover, these are not just matters of concern for the discipline of sociology; the forms of knowledge which we develop have wider political implications. Carl May puts this well:

619

The state in contemporary Britain is increasingly characterized by new kinds of reflexivity, mediated through systems and institutions of technical expertise – in which policy rooted in *evidence* is central to its strategic practices, and thus to political discourse. These are expressed in many ways, but involve a central shift towards the primacy of (largely quantitative) knowledge as the foundation for an increasingly active and managerial model of state intervention across a range of policy fields. The emergence of this imperative towards *evidence-based* policy . . . is one important ideological feature of the apparently post-ideological character of contemporary British politics. In the British case this has involved the rapid development of policy mechanisms and agencies through which this work can be effectively delegated to the Academy. [. . .] One outcome of this is that sociologists might now find themselves among the outsourced civil servants of the evidence based state. This is why political contests about methods are important. (May, 2005: 526–7)

So what, ultimately, do we want our political contests about methods to be about? As far as understanding where British quantitative sociology is today, we have suggested that the shift from 'causality to description' suggested in the title of this paper is too simple – hence the question mark. As we hope to have shown, the 'descriptive turn' is partly the result of inscriptive devices, such as SPSS, which have assisted in its emergence and propagation.

As powerful global actors increasingly come to act on the outcomes of 'predictive analytics', sociologists – who have, after all, contributed to their emergence – need to reassess their standpoint. They not only need to 'get inside' the technology in order to report back on its functioning, they also need to emulate this particular inscription device in order to generate alternative readings of data already subject to analysis and, crucially, offer parallel reportage on data that, although perhaps of no interest to commerce, can offer some enduring sociological insights.

Notes

1 For those who are less familiar with SPSS, visiting this website may still be worth doing just to see how we might imagine qualitative software and text mining technologies evolving in the future.
2 Available online at http://www.spss.com/corpinfo/history.htm (accessed 17 June 2008).
3 Thanks as well to Aidan Kelly, Andy Tudor and Jonathan Bradshaw for offering their recollections.
4 The second named author was fortunate enough to work on part of this project when still an undergraduate student at the University of Surrey in 1982. He helped produce an SPSS teaching guide on 'Class and Stratification' using the GHS data 1973–1982 and a Prime Mainframe computer. It was the first time he had used a word processor – a package called Wordstar.
5 The final named author kept his daughter's primary school supplied with rough sketch paper for several years by passing on hundreds of pages of output on a monthly basis!

6 Many of the texts published in the *Contemporary Social Research* series edited by Martin Bulmer may also be included here.

7 One reason for this, of course, is that such procedures are not as yet part of the standard SPSS product familiar to British academics. They are costly and require an additional licence.

8 Introduced, as a pen-and-paper technique, at a state secondary school to the 13 year old son of the second named author during the initial drafting of this paper in 2008.

References

Abrams, P., (ed.) (1968), *The origins of British Sociology 1834–1914*, Chicago: University of Chicago Press.

Beer, D. and Burrows, R., (2007), 'Sociology and, of and in Web 2.0: Some Initial Considerations' *Sociological Research Online*, 12 (5): ⟨http://www.socresonline.org.uk/12/5/17.html⟩.

Bradbury, M., (1975), *The History Man*, London: Picador.

Bryman, A. and Cramer, D., (2008), *Quantitative Data Analysis with SPSS 14, 15 and 16: A guide for social scientists*, London: Routledge.

Bulmer, M., (ed.) (1985), *Essays on the History of British Sociological Research*, Cambridge: Cambridge University Press.

Burrows, R. and Gane, N., (2006), 'Geodemographics, Software and Class', *Sociology*, 40 (5): 793–812.

Byrne, D. and Ragin, C., (forthcoming), 'Handbook of Case Based Methods', London: Sage.

Byrne, D., (1998), *Complexity Theory and the Social Sciences*, London: Routledge.

Byrne, D., (2002), *Understanding Quantitative Data*, London: Sage.

Cartwright, N., (2007), *Hunting Causes and Using Them*, Cambridge: Cambridge University Press.

Dale, A., Arber, S. and Procter, M., (1988), *Doing Secondary Analysis*, London: Unwin Hyman.

Field, A., (2007), *Discovering Statistics using SPSS* (2nd edition), London: Sage.

Fielding, J. and Gilbert, N., (2006) *Understanding Social Statistics*, (2nd edition), London: Sage.

Fielding N. and Lee R., (1998) *Computer Analysis and Qualitative Research*, London: Sage.

Geertz, C., (1973), *The Interpretation of Culture*, New York: Basic Books.

Gilbert, N., Dale, A. and Arber, S., (1983), 'The General Household Survey as a Source for Secondary Analysis', *Sociology*, 17 (2): 255–259.

Gilbert, N., (2008), *Agent Based Models*, London: Sage.

Gilbert, N. and Troitzsch, K., (2005), *Simulation for the Social Scientist*, (2nd edition), Buckingham: Open University Press.

Graham, S., (2004), 'The Software-Sorted City: Rethinking the "digital divide"', in Graham, S., (ed.), *The Cybercities Reader*, London: Routledge: 324–332.

Graham, S., (2005), 'Software-sorted geographies' *Progress in Human Geography*, 29 (5): 562–580.

Grier, D.A., (2005), *When Computers Were Human*, Princeton: Princeton University Press.

Hindess, B., (1973), *The Use of Official Statistics in Sociology*, London: Macmillan.

Irvine, J., Miles, I. and Evans, J., (eds) (1979), *Demystifying Social Statistics*, London: Pluto Press.

Kent, R., (1981), *The History of British Empirical Sociology*, Aldershot, Hants: Gower.

Law, J., (2003), *After Method: Mess in Social Science Research*, London: Routledge.

Lyon, D., (2002), 'Surveillance studies: understanding visibility, mobility and the phentic fix', *Surveillance and Society*, 1 (1): [Online] http://www.surveillance-and-society.org.

Lyon, D., (2003), *Surveillance as Social Sorting: Privacy, Risk and Digital Discrimination*, London: Routledge.

Marsh, C., (1989), *Exploring Data*, London: Sage.

May, C., (2005), 'Methodological Pluralism, British Sociology and the Evidence-based State: A Reply to Payne *et al.*' *Sociology*, 39 (3): 519–528.

McCue, C., (2007), *Data Mining and Predictive Analysis: Intelligence Gathering and Discovery*, Burlington, MA: Butterworth-Heinemann.

Pallant, J., (2007), *SPSS Survival Manual: A step-by-step guide to data analysis using SPSS version 15*, (3rd edition), Milton Keynes: Open University Press.

Parker, S., Uprichard, E. and Burrows, R., (2007), 'Class Places and Place Classes: Geodemograph- ics and the Spatialization of Class', *Information, Communication and Society*, 11 (6): 901–920. Payne, G. and Williams, M., (2005), 'Generalization in Qualitative Research', *Sociology*, 39 (2): 295–314.

Phillips, D. and Curry, M., (2002), 'Privacy and the Phenetic Urge: Geodemographics and the Changing Spatiality of Local Practice', in *Surveillance as Social Sorting: Privacy, Risk and Digital Discrimination*, Lyon, D., (ed.), London: Routledge: 137–152.

Ragin, C., (1987), *The Comparative Method: Moving Beyond Qualitative and Quantitative Strate- gies*, Berkeley, CA: University of California Press.

Ragin, C., (2000), *Fuzzy-Set Social Science*, Chicago, IL: University of Chicago Press.

Richards, L., (1999), 'Data Alive! The Thinking behind NVivo', *Qualitative Health Research*, 9 (3): 412–428.

Richards, T., (2002), 'An Intellectual History of NUD*IST and NVivo', *International Journal of Social Research Methodology*, 5 (3): 199–214.

Savage, M., Bagnall, G. and Longhurst, B., (2005), *Globalization and Belonging*, London: Sage.

Savage, M. and Burrows, R., (2007), 'The Coming Crisis of Empirical Sociology', *Sociology*, 41 (5): 885–900.

Thrift, N., (2005), *Knowing Capitalism*, London: Sage.

Tufte, E., (1983), *The Visual Display of Quantitative Information*, Cheshire, CT: Graphics Press.

Tufte, E., (1991), *Envisioning Information*, Cheshire, CT: Graphics Press.

Tufte, R., (1997), *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, CT: Graphics Press.

Tukey, J., (1977), *Exploratory Data Analysis*, Menlo Park, CA: Addison-Wesley.

Urry, J., (2003), *Global Complexity*, Cambridge: Polity Press.

Veldman, D., (1967), *Fortran Programming for the Behavioral Sciences*. New York, Holt, Rinehart and Winston.