

Statistics as Reification

Simeon Scott

s.scott@bradfordcollege.ac.uk

Introduction

This text addresses the major political and philosophical issues in statistics and data science. The text's main arguments arise from outside of the statistical paradigm, with what could be called a *Critique*. Thus, the thrust of the text is criticism with a view to transforming capitalist social relations. This approach contrasts sharply with that offered by Huff (1976) and his contemporary disciples, who present criticisms of statistics from within its own paradigm. The text begins by pointing to the avalanche of statistics typically presented to citizens in what one writer has called *the spectacle*. There follows a discussion of those actually producing statistics, including a social class perspective ranging from prehistory to the present day. Given the tarnished history outlined in this section, the text moves on to address the issue of whether, content aside, the techniques used by statisticians are *neutral*. To this end, the text examines the mean value, showing how this artifact has profoundly affected the lives of working class people. Noting the Cambridge Analytica fiasco, the article then considers the big data phenomenon, including its big brother implications. Prior to the conclusion, there is a short discussion of econometrics. Let us then begin with the data avalanche.

Another second another statistic

In Britain, as elsewhere, there are currently a large number of organisations in both the corporate sector and public sector routinely collecting and analysing data. As they gradually replace a shrinking welfare state provision, charities too are increasingly involved in gathering data. Much of this data is summarized and published, with the broadcast, print and other media offering us a daily diet of figures: Marks & Spencer's profits have fallen by $x\%$, inflation is $y\%$, unemployment is $z\%$ and so on. In fact, it is no exaggeration to say that, around the

globe, in such areas as economics, crime, physics, chemistry, medical science and indeed virtually every area of human endeavor, a mass of data is collected every second of every day. Debord (1977) used the term *spectacle* to refer to the corporate media, screen images, communication technologies, advertising hoardings and more that typified 1960s capitalism. This spectacle, he argued, turns authentic human life upside down, or “inverts” it, by reducing people to passive observers rather than active participants in day to day life. It is clear that in the early 21st century this state of affairs has become both quantitatively and qualitatively more intense, due to this mass of numbers and the near ubiquity of the smart phone.

In the workplace, given the pervasiveness of computer hardware, software provides management information systems that gather data, overtly and covertly, to monitor the performance of wage workers. Apologists for this state of affairs take the view that the statistical techniques applied to this data are tools to make government, business and other agencies more efficient, with politicians and managers relying on objective facts, rather than anecdote or prejudice, when making decisions. Yet, such arguments in favour of quantitative expertise so as to facilitate business efficiency and ‘western liberal democracy’ have been challenged by those with other perspectives. One counter argument is that in a capitalist society the statistical process is inherently political, tending to give authority to the values and interests of the dominant class, which tends to monopolise decisions regarding the resources necessary to facilitate statistical projects. With reference to the oft-mentioned term *data-driven decision making*, critics argue this is merely a euphemism for ensuring that the ‘correct’ decisions are made by the ‘correct’ people.

Lies, damn lies and statistics

A common response to this flood of statistics and the uses to which they are put, is to reiterate the arguments present by Huff (1976) and his contemporary supporters. They argue much of the output of statisticians, or data scientists as they are increasingly called, has, by accident or design, the effect of fostering confusion or manipulation. The gist of this view is that, either due to a lack of technical competence or the desire to persuade, much of the output of statistics is at best half-truths or, in some cases, downright lies. Simply put, the argument is

that much of descriptive and inferential statistics does not adequately correspond to the reality to which it refers. Revealing gross sociopolitical naivety, a commonplace amongst statisticians, most of Huff's disciples have no problem with current statistical techniques, the ubiquity of data or the global capitalist social relations to which they refer. Rather, they argue that there are statistics that are done right and statistics that are done wrong, with the former offering objectivity and the latter either lacking competence or intended to deceive. In summary, this is a criticism from inside of the statistics paradigm.

Number and social class: who produces statistics?

The editors of Dorling and Simpson (1999), along with other "left wing" statisticians, highlight the socioeconomic and political uses of statistics. However, they reject the argument that "every statistic" should be dismissed "because it inevitably turns people into objects to be manipulated and controlled". They correctly reject the view that statistics is "ethically neutral, asocial", producing "bodies of knowledge and techniques...used for good purposes or abused for bad ones". Yet, Dorling and Simpson appear to want more statistics, rather than less, arguing that: "With a little thought and a little knowledge you can produce statistics which are fair and just" (415 and 419). Paraphrasing Mandy Rice-Davies writing about career politicians, we could note that career statisticians would say that wouldn't they. Whilst the argument that statistics should become democratized and then it will become "fair and just" has merit, unless this forms just one part of a fundamental social transformation, then it is merely a one-sided truth as Hegel would say. This discussion begs a number of questions; let us begin with: who produces this ever increasing mass of data?

Rejecting the claim of quantitative objectivity, it will be argued here that not only the content, but also the form, of statistics has been profoundly affected by the social context and ideology of the social class that typically produces them. In order to provide evidence for this argument, we can briefly review the history of quantification in general and statistics in particular. Anthropological research would suggest that in prehistory, whilst there was much variation, the use of numbers was more or less restricted to the role of adjective or predicate in practical contexts: two rabbits, three trees and the like; with no particular social group monopolizing the use of number. Hunter-gatherers

and early agriculturalists used quantification as adjectives for objects with similar qualities, with the limits of this process being determined by such practical considerations as, for example, when an animal was ill or young it would not be equivalent to an adult healthy one. Following the agricultural revolution and the rise of the ancient city states, there developed a class of initiates, including Egyptian priests and Babylonian astronomers, who supported the interests of their respective ruling elites. Whilst this intermediate class used quantification for a range of practical purposes, the records show that abstract mathematics developed and, as Hersh (1997) argues, numbers became nouns or subjects in their own right. There remains a Eurocentric tendency to lionise all things Greek; the claim being that these ancients invented mathematics, philosophy, democracy and much else. It is true that there existed Greek groups of initiates, the best known of which were the Pythagoreans. However, thanks to the research of African-American academics and others, including Bernal (1987), it is now clear that the ancient Greeks themselves tell us of their travels to Egypt and elsewhere to study mathematics and much else. With regard to the ancient city states of the Tigris Euphrates basin, Levy (2013, 14) writes of tallies used to record “taxes, tithes, census data, dates, land” with an intermediate class of bureaucrats, astronomer priests and proto-accountants using clay tablets to record “weights and measures, squares and cubes and reciprocals...(and) compound interest” (35) in the service of the ruling elite.

With regard to nascent statistics, the feudal records show that data was collected by a relatively well rewarded group of employees of the ruling class. A census would give the feudal monarch data that would be useful for such purposes as inventory, taxation and military service. As trained statisticians know, modern sampling techniques rely heavily on probability theory, which was developed by Cardano and de Fermat in the 16th century to assist members of the European aristocracy with a penchant for gambling. As probability theory became formalized it was taken up by joint stock company investors who realised that profits could be made in a range of activities such as insuring ships engaged in the slave trade and various colonial enterprises. In the late 18th early 19th century, the Belgian academic/astronomer Quetelet turned his attention to social statistics and was to be highly influential on the wealthy English gentleman Francis Galton. Macken-

zie (1981) has argued that, as is the case today, in the Victorian period the social class that produced statistics was the professional middle class¹. Whilst this is largely true, Galton was from the capitalist class, the family fortune coming from banking and arms manufacturing.

A professional middle class emerged in the European colonial period and established itself during the industrial revolution. Like accountants, lawyers, architects and others in their respective areas of competence, career statisticians came to monopolise the collection and analysis of data. Then, as now, statisticians tended to be financially comfortable without being rich, caught between the capitalist class and the working class. In order to maintain their relatively privileged life style, statisticians tended to propagate the interests of capital. However, members of this class often took the view that those ruling the major European states, including the remaining members of the aristocracy and the new industrialists, had by and large obtained their wealth and power either by inheritance or good fortune given the swings and roundabouts of corporate profitability and bankruptcy. Therefore, the will to power of this section of the intelligentsia was predicated on their belief in genetic heritability, or eugenics, specifically the view that their academic achievements were the result of their intellectual superiority over other classes as measured by their intelligence quotient (IQ). Such a view was propagated in the private schools and elite universities of late 19th and early 20th century Europe and North America². In Britain, for the Victorian statistical intelligentsia, who alleged that their mental prowess entitled them to a relatively privileged status, Pythagoras, Gauss and others became archetypal mathematical “geniuses” and models for their intellectually challenging endeavors.

Bolshevism and statistics

This intellectual elitism was taken up by the middle class members of the Bolshevik Central Committee in the early 20th century. Seeking to enroll support amongst the Russian working class, the Bolsheviks re-wrote history by claiming that they were “Marxists”. In *Das Kapital*, Marx (1976) showed himself to be at ease using data taken from a

¹ Bowlby was an exception – see attached, appearing in this issue

² Not true for Bowlby and Stephenson

range of “bourgeois” sources, including factory inspectors’ reports. However, Marx used these statistics as just one side of his analysis of capitalism, the primary aim of which was to show how the *internal contradictions* generated by the wage labour system could not be overcome within the framework of capitalism itself. Citing the achievements of the German worker/philosopher Joseph Dietzgen, and despite his middle class origins, Marx rejected the view that working class people were mentally inferior to middle class intellectuals. Marx was also no supporter of either state capitalism or rule by a “vanguard” in a one party state. Yet, the leading Bolshevik Lenin embraced these dogmas with a passion. Whilst not explicitly embracing eugenics, as trained lawyers, Lenin and Trotsky were a paradigm case of the ambition of the radical middle class. Angered by the marginal status offered to them in Tsarist Russia, the Bolsheviks saw events in 1917 as an opportunity to introduce state capitalism into the country with themselves as its architects and guardians. The Bolshevik coup had been supported by large sections of the working class, including units in the army and navy. However, by 1921 it became clear that large sections of the working class and peasantry wanted to press forward with revolution and rejected the Bolshevik one party state agenda. Well aware that they would be sidelined, if not arrested for crimes against the working class, Lenin and Trotsky employed extreme violence to crush all working class and landless peasant opposition to their state capitalist plans, with independent trade unions, soviets and workers’ councils eliminated.

Despite having no training in statistics, Lenin presumed to comment on factory statistics published during the Tsarist period: they remained, he argued, “completely obsolete and inefficient in terms of quality as well as structure”; quoted in Suhara (2017, 3). Having achieved his goal of absolute power following the civil war, Lenin used the relatively well equipped statistical agencies to produce data to justify his state capitalist programme. Having realised the opportunities offered to professionals by state capitalism, some British intellectuals, including mathematicians, rushed to join their local Communist Party. Meanwhile, Lenin centralised the statistical agencies and insured that the data collected was adjusted for propaganda purposes. Famously, Lenin used underestimated figures on agricultural output to justify his New Economic Policy, following the blood soaked suppression of work-

ers' opposition at Kronstadt, Tambov, the Ukraine and elsewhere. Introducing Taylorist time and motion practices, piece rate incentives and other productivity initiatives pioneered by Taylor in the United States, as discussed below, Lenin showed no inclination to publish statistics on the rise in industrial injuries, strikes, go-slows and outright sabotage, all of which became commonplace in the early 1920s and beyond.

Apparently more modestly intellectually gifted, as the son of a shoemaker, Joseph Stalin sacked, exiled or executed those statisticians who dared to challenge the ideologically driven data produced by the Bolshevik state. Thus began the simultaneously tragic and laughable output of fake data on grain and tractor production, to name but two examples, intended to hide the extent of the various famines that were the direct result of the Bolshevik collective farm programme. Despite all the false or missing data, which was justifiably mocked in the western media, it seems that by the 1960s only in rocket production was the USSR able to keep up with its western rivals. In a document released by the CIA (2011), the journal *Vestnik statistika* reports on a 1950 conference at which "Marxist-Leninist" delegates complain about "bourgeois" statistical theory that is "in conflict with the tasks proposed for Soviet statistics by the Communist Party of the Soviet government...The tasks of statistics and the theoretical foundation of statistical methods are formulated in the works of Lenin and Stalin...with the purpose of building a Communist society...Only statistics based on historical materialism and political economy is scientific statistics" (1-2). Criticising statistical theory based on probability and the law of large numbers, the orthodox delegates argued this is a bourgeois view "having nothing in common with Marxism...Soviet statisticians should be educated in a spirit of unlimited devotion to the Communist Party headed by the great Stalin, and improve practical Soviet statistics along the lines indicated by the Party and the government" (3). The drift of the conference debate was that "bourgeois" modes of statistics were alright when applied to natural phenomena, but need to be adjusted to the "dialectical" approach of "Marxism-Leninism" when applied to social being.

IBM, statistics and the Nazis

From this low point in the history of statistics things were to get worse, much worse. Given the strong connection between eugenics on the development of Victorian statistical theory, the status attached to high IQ and the respectability of Eurocentric racial theory, it is no great surprise to learn that a number of statisticians brought deep shame on their profession. As Black (2004) explains, there is overwhelming evidence that statisticians working for IBM provided equipment and ‘expertise’ to facilitate the Holocaust in the 1930s and ‘40s. Central to this outrage was the punch-card machine, then state of the art specialist data gathering equipment, which was made available to IBM’s subsidiaries in Romania, Switzerland and Germany. In 1941, reports Black, with the full knowledge of the then IBM president Thomas J. Watson Sr. located in the New York head office, the Romanian Central Institute for Statistics used this census orientated equipment to systematise the extermination of named members of particular racial, ethnic, religious and political groups. IBM’s machines were used by the pro-Nazi Romanian leadership to facilitate the arrest and transportation of targeted individuals to the death camps. In the camps, reports Black, the IBM code for Jews was 8, for Gypsies 12, for general executions 4, for death in gas chambers was 6. Each concentration camp had its own IBM code: Mauthausen 7, Buchenwald 2 and Auschwitz 1.

Statistics and identity politics

Following this low point in the ideological genesis of career statisticians, who perhaps argued they were “just carrying out orders”, members of this profession have for the most part abandoned eugenics, although as contemporary textbooks and journal articles show, many remain preoccupied with IQ tests and Eurocentric views on mathematics. Following World War II, most government statisticians provided data in support of the Keynesian welfare state consensus, but in the 1980s were required to gather data in support of the new neoliberal agenda. With the expansion of higher education in the 1960s, some people of working class origin, along with middle class students exposed to Marxism in their university unions, became career statisticians. Typical of this trend are the authors included in Irvine et al (1981) and Dorling and Simpson (1999), most of whom would probably describe themselves as “left-wing”. Yet, this spatial metaphor tends

towards mystification rather than enlightenment, since these statisticians have for the most part embraced the politics of identity.

Propagated by former UK Prime Minister David Cameron, in order to widen the electoral appeal of the Conservative Party, the politics of identity has become a substitute for fundamental social change for many on “the left”. In recent decades working class people have had imposed upon them, with little or no democratic debate, a lexicon of terms along with workplace rules and regulations. Whilst there is much that is progressive in the UK’s Equality Act (2010), such legislation and the cultural norms it encourages are most notable for what they exclude. Whilst there are statistics on low pay, zero hours contracts and rising inequality; there are few statistics that are presented in a way that would be helpful in establishing democratic workplace decision making, justifying radical salary reductions for CEOs and other policies that would challenge the global wage labour system. By offering identity politics as a gloss on a neoliberal philosophy that has held sway since the early 1980s, statisticians are content to collect data on safe statistical categories, such as the number of women or BME people that are directors of FT100 companies. However, this denies the validity of, and in effect the existence of, those categories which present a fundamental threat to the neoliberal socioeconomic agenda.

Numbers: are they neutral?

Given this ‘dramatic’ history of career statisticians, apologists for the profession argue that despite its applications, be they good or evil, the techniques used in contemporary statistics are *neutral*. Although few of them reflect on the philosophical or socioeconomic aspects of their craft, by default most statisticians take the view that, like quantification in general, statistical techniques, such as the method of calculating a mean value, are not *invented* but *discovered* in some timeless world of the *Forms*, as Plato called them. Yet, as the humanist mathematician Hersh (1997) argues, numerical artifacts are not discovered in some other realm, but rather have had a long socioeconomic genesis. The Victorian statisticians played a key role in the relegation of quality and the promotion of abstract quantity; their ideas on social class, race and inherited intelligence being sublimations of their ambivalent position between capital and labour in a world of colonial conquest and the triumph of industrial capitalists over landed aristo-

crats. In short, different socioeconomic circumstances will tend to promote different statistical artifacts. We can only speculate on how statistical practice, and metrification more generally, would have developed had historical circumstances been different. Rather than $1+1$ always being equal to 2, as ordained by Plato's world of the Forms, Hersh shows that addition, not to mention the mass of other mathematical operations, is no routine procedure that can be successfully applied at will. $1+1=2$ is a cultural artifact predicated on the need to facilitate the trillions of daily transactions that are the life blood of the capitalist mode of production. Hersh shows the limits of this artifact, giving a number of examples of when it does not apply. Most schoolchildren around the globe are more or less forced to learn arithmetic; with addition, subtraction, multiplication and division seen as essential to functioning in their allotted role in capitalist society. Along with profit and loss calculations and other accounting techniques, in school textbooks arithmetical skills are typically applied to calculating VAT, revenue maximisation and a range of related 'skills'.

However, despite their practical role in serving their lords and masters, the intelligentsia in ancient and medieval times tended to view numbers in a radically different way than is the case in contemporary capitalist societies. Prior to the scientific revolution of the 17th and 18th centuries, for members of the intelligentsia, such as John Dee, numbers were not merely abstract measures of quantitative difference as is the case today, but rather they were part of a religious, mystical, astrological and alchemic whole. As Rogerson (2013) indicates, prior to this revolution, each number was *qualitatively* different from all others; for example the number 7 was given great significance in terms of the planets, the diatonic music scale, the days of the week, alchemic experiments and more. The number zero, which is crucial in contemporary statistics, was invented in ancient India; it was introduced into Europe, via the Muslim world, by merchants and accountants. Despite strong resistance from the Catholic Church, which abhorred the notion of nothing, following its widespread use in business as a placeholder in the base 10 number system, it became an abstract symbol for no thing. However, whilst nothing can refer to such qualitative states as starvation or bereavement, given its quantitative genesis in capitalism, zero has been robbed of these associations. The number 13's Satanic associations, along with other remnants of qualitative notions of number, are today dismissed as superstition and 13 has be-

come just another abstract number; although in some cultures its demonic heritage survives to this day. During the scientific revolution abstract Arabic numbers, robbed of their qualitative significance, became part of a system of investigation and experimentation in which, for the most part, nature was to be broken up into its component parts and dominated as part of colonial expansion. Thus numbers metamorphosed into abstract nouns and became the quantitative lubricant of the capitalist mode of exchange predicated on the idea of measurable price and quantity. This process is described by Hegel, who refers to the analogy of the adjective red which was a description of the noun rose. In becoming detached from this noun and itself turning into the noun red, argued Hegel, we have lost the beauty of the rose and are left with an abstract colour.

The mean value as reification

Debord's (1977) earlier mentioned text brings to mind the word *reification*, which comes from the German term *Verdinglichung* as used by the philosopher Hegel. The word came to be used in English during the industrial revolution in the mid-19th century, along with related terms such as *alienation*, *fetishism* and *commodification*. Reification refers to processes that are typical of capitalism whereby workers, consumers and others become dehumanised. The word literally means that human relationships are turned into *things*, which in turn take on a life of their own. Thanks to the Victorian pioneers, the average or *mean* value has become the cornerstone of contemporary statistics, with correlation, regression, hypothesis testing and more all predicated upon it. Although making no mention of reification, and revealing a socioeconomic naivety typical of statisticians, Rose (2015, 45) nevertheless offers a good description of how the mean value has become a *thing* that appears to control the lives of most workers. He describes his work, involving following a script, in a call center which is worth quoting in full: "Since following the script correctly meant that a customer service call would last an average length of time, I was evaluated on the duration of each and every call. If a call exceeded the average time, my screen began flashing red. Instead of focusing on the quality of the call, I focused on making sure I hit the disconnect button as fast as possible. The computer updated my average time after each call and showed how I compared to the group average – and shared my average with my supervisor, too. If my average exceeded the group average by too much, my supervisor paid me a visit, which he did,

several times. If my average had remained high, he could have fired me – though I quit before that could actually happen”. This reified *appearance* is true and very real; however, we have to understand how this state of affairs has come into being: what are the *essential* relationships at the core of this process?

Whilst there has been much discussion of how Quetelet’s view of deviations from the mean value, which he argued were “errors”, differed from that of Galton. The latter argued deviations were indicative of “rank” and, in terms of intelligence, Galton argued those markedly below the mean or “Mediocrity”, were “Imbeciles” and those markedly above were “the Eminent”. What marks Rose’s text is that he takes this debate, including the contradictions between the mean as the *ideal*, as per Quetelet, and mean as mere *mediocrity*, as per Galton, and shows how the management theorist F.W. Taylor and later the psychologist Thorndike interpret this contradiction in the context of both wage labour and the capitalist education system. Quoting Taylor, Rose (42) writes: “In the past the man was first, in the future the system must be first”. Thus began the use of the mean in factories across the globe including, as we have seen state capitalist nations such as the USSR, as a standard for the time and motion of wage labour. Whilst the mean value would tend to change over time, in order to increase labour productivity, it remained the basis for the division of labour. The new “managers”, as Taylor called them, were to ensure that deviations above or below the mean were minimized so as to maintain coordination in the work process. “(E)ach worker”, writes Rose (47), became “like a cell on a spreadsheet...a number in a column, and interchangeable Average Man”.

The western education system was similarly refocused so as to facilitate Taylorism; with Mencken arguing “The aim of public education is not to spread enlightenment to all; it is simply to reduce as many individuals as possible to the same safe level, to breed and train a standardized citizenry, to put down dissent and originality”; quoted in Rose (51-2). Thus Galton’s view of the mean was resurrected, with education becoming a process whereby citizens were appointed to their proper station in life: managers or workers. This process was legitimised by Thorndike, a big fan of Galton, who differentiated between the dull and the talented on the basis of their deviation from the

mean, the latter being “the standard”. To this day, following Galton and Thorndike, the mean as a “standard” of mediocrity underpins assessment in capitalist education from the elite universities, at one end of the spectrum, to special needs schools, at the other.

Young (1981) offers the example of the manufacture of a chair to demonstrate another dimension of the reification process. Previously such a commodity would have been produced by a craftsman, who would have combined quantity and quality in a personal relationship with its users. Today, such personal relationships are the exception; typically the production of a chair abstracts from the personal and qualitative; abstract numbers being the lubricant for a division of labour that includes the chair’s design, marketing, quality control, accounting and more. However, here *quality control* has become a euphemism for statistical testing so as to provide a minimal standard which, through guarantees and warranties, in turn provides the opportunity for profit making. Similarly, in contemporary capitalist society, health is a euphemism for fitness for work, which in turn offers the potential for private profit. The UK’s National Health Service is systematically being privatised, with ever more management consultants, statisticians and others offering reified proxies, such as BMI based on deviations from the mean, instead of supporting the general wellbeing of the population.

Finally in this section on reification, we can note that for most professional statisticians the world consists of apparently self-subsisting objects that can be readily quantified and thus become *data* or *facts*. Mystifying what are in reality the social *relationships* of capitalism, we can note the existence of three mean values, 36 24 36, to describe the ideal body shape of a woman. This reification abstracts from a mass of fundamental issues regarding gender relations in a capitalist society. Similarly, a newspaper might release a statistic telling its readers that a company pays an average wage of £30,000 per annum. This reduces human labour to its appearance in capitalist social relations: a burdensome cost which is a deduction from corporate profits. Yet, in its essence such a figure should be the starting point for a study of the contradictions inherent in the polarity between workers, for whom wages are their source of sustenance, on the one side, and capital, on the other. In legitimizing themselves by association with Plato’s ideal, for the most part statisticians serve their capitalist masters by obscur-

ing the qualitative essence of working class life, thus undermining their striving for autonomy and the will of workers to change society. Notwithstanding his socioeconomic naivety, Muller (2018) shows how human creativity is manipulated and stifled by the metrification of education, health and the wider wage labour system. Using the term “data arms race” (10), he shows how “targets” or “benchmarks” are set, typically using the mean, in order to reduce costs; arguing that these processes create “conflict” and “declining morale” (9). In fact, this process of metrification has become the battleground on which the internal contradictions of the wage labour system are fought out.

Big data: the end of statistics?

Mayer-Schönerberger and Cukier (2017), with an unwitting nod to Hegel, refer to the effects of using big data rather than relatively small samples from a population. They write: “the quantitative change has produced a qualitative change...by changing the amount, we change the essence” (10). However this may be, in recent years the relatively cosy world of academic statisticians has been undermined by developments in big data collection and analysis. The near ubiquity of the computer, along with the privatisation of the internet in 1990s, has given rise to the dominance of a few internet-based companies. With the largest turnovers in the world, these companies make profits by collecting and selling data from internet users to facilitate targeted advertising and other forms of marketing. Given the potential incomes and bonuses available in the new profession of data scientist, many potential statisticians are having second thoughts about a career in the academy or the civil service. One writer who is well disposed to data science, Clegg (2017, 118), describes himself as having “a skill that is relatively scarce”; he repeats the Platonic myth that big data is “neutral...It can’t do anything on its own” (139). Telling us much about the corporate-orientated ideology of this new breed of data professionals, Clegg argues that, via his smartphone and other devices, he routinely uses the services of Uber, Starbucks, Google, Apple and others “to make my life easier” (47). That this author is untroubled by big data companies influencing elections, referenda and the representative democracy process in general is clear when he writes: “at the moment Facebook is ethical and unbiased. But let’s imagine that in the future the company was bought by a malignant power” (84).

Clegg offers little more than a static cost-benefit analysis of data science, with all the good aspects on one side and the bad aspects on the other. He offers some minor criticisms concerning the use of big data driven algorithms as used by employers and what he calls the “big brother” problem. A number of writers have bemoaned the eclipse of statistics in the light of technology driven *data mining*, i.e. the collection of large data sets, including $n=all$, which eliminates the need for sampling and the range of techniques associated with it, thus presenting a threat to the more traditional craft of the professional statistician. Davies (*The Guardian*, 19th January 2017), for instance, argues the “populist right” is anti-statistics; claiming “a new age of big data controlled by private companies is taking over – and putting democracy in peril”. In fact the black boxes that hide the algorithms typical of the big data revolution use the same statistical techniques as presented in the traditional textbooks, mean values, correlation, regression and the like. However, notwithstanding the use of these techniques in the service of mass marketing by the world’s large corporations, there is an important issue raised by the heroic Edward Snowden.

The data scientist as big brother

First published in *The Guardian* (6th June 2013), former US intelligence officer Edward Snowden reported on his work as a data science contractor for the National Security Agency (NSA). Snowden revealed that we should not only be concerned about the data collection of private “tech” companies. This courageous whistleblower informed the world that American government agencies were training this new breed of “scientists” to collect and analyse both data and metadata, the latter referring to communication facts and figures rather the contents of these communications, on their own citizens, a process they justified as part of their antiterrorist activities following 9/11. Most, if not all, of the tech giants have participated in these spying programmes, Snowden reported, whilst publicly claiming that they did not do so. As Harding (2014) reports, the implications of Snowden’s revelations was that these programmes could, and indeed were, being used for surveillance way beyond the limits of what was required to keep Americans safe. Simply put, Snowden revealed that US security agencies were breaking the law, specifically the fourth amendment, by spying on citizens by collecting their data and their contacts without either their consent or good cause. To this end, Snowden revealed,

these agencies were both using secret court hearings and forcing telecoms companies and others to participate in such spying.

In Britain things are even worse, as Snowden revealed that, with the permission of Prime Minister Gordon Brown and Foreign Secretary David Miliband, GCHQ was doing the same things, but with no constitutional or legal impediments to limit their activities. On the issue of data encryption, which was suggested as a means of protection against government, Snowden pointed out that most of the commercially available systems had had their data encryption keys hacked by the security agencies. In short, in answer to the question: who collects big data from the internet? Snowden's answer is the NSA does in cahoots with the commercial giants. He reports that the NSA, which is the largest US employer of qualified mathematicians, targets people who seek to adjust their privacy settings. Following Snowden's revelations, the author of the text before the reader had trouble convincing members of Bradford Momentum, a local branch of a national support group for Jeremy Corbyn, which organises its activities via the internet and includes a trained statistician, that it is highly likely that GCHQ routinely collects data on Momentum's activities.

The standard government/corporate response to Snowden's leaks is that if a person is not doing anything wrong they have nothing to fear. In answer to this several writers, notably Schneier (2016), have argued that a person's political affiliation, particularly if it is not part of the so-called center-left center-right waveband, is of great interest to governments. Data is liable to be used for any purpose deemed fit by governments and corporations with little or no democratic accountability. According to Schneier, there are certainly 17, and maybe 18, American intelligence agencies collecting data and it is a moot point, he adds, whether most this is legal according to US law or, in the case of GCHQ, British law. Schneier confirms all nations, i.e. those with the resources to do so, hack each other's networks and install malware for both military and commercial reasons. However, he reiterates the point that one cannot clearly distinguish between state and commercial monitoring because they are intertwined. Often the courts, sometimes sitting in private and secret, give permission for such sharing.

EU data protection laws, which are stronger than those elsewhere, claims Schneier, are evaded by all manner of data collection organisa-

tions. He argues such gross invasions of privacy are a breach of human rights intended, and often succeeding, in creating political, social and cultural conformity, obedience and submissiveness. We become less likely to stand up for our beliefs through civil disobedience, demonstrations, strikes and the like. The EU's General Data Protection Regulation (GDPR), which is intended "to return to citizens control of their personal data", takes effect in May 2018. However, most of the EU's decisions are made by unelected commissioners, whose record on resisting corporate lobbying is, at best, poor. Readers may be aware of an alliance of governments and companies entitled ID2020 which plans to introduce "a unique digital identity for everyone on the planet"; Ward (2018)³. Noting dubious claims that this plan is a "humanitarian cause...to unify the world's citizens participation in a global community...with a person from birth to death", Ward asks some pertinent questions including: what if you do not want to participate? and who will administer this identity system?

As Schneier confirms, anonymous data bases can easily be de-anonymised with only a few correlations; and other attempts to hide data from corporate surveillance is met with ever more sophisticated statistical techniques to outwit these attempts. With regard to data storage, much of this is outsourced and retained on large cloud computing servers. Schneier points out that corporations store our data on servers located in countries which have less rigorous laws, if any at all, on data protection, de-anonymisation and the like, than those in Europe or North America. Turning to predictions based on big data correlation and regression algorithms, including crimes yet to be committed, these typically fail to address the issue of the correlations between social class and crime. So, for example, higher income people tend to commit fraud, often a grand scale, insider trading or tax evasion/avoidance which are often not treated as crimes, particularly in the UK. In contrast to this, lower income workers are more likely to get involved in shoplifting or burglary, which are treated as crimes, although cuts in police numbers may well mean that few resources are put into investigating them. Either way, few data scientists express an interest in the kind of social change that might lessen both actual and predicted crime. Notwithstanding the Cambridge Analytica fiasco, preventing big data gathering along with its sale and purchase by other

³ The problem is they are missing at least 700 million! (see Carr-Hill, 2017).

parties, without the consent or even the knowledge of internet users and others, would undermine the staggering profits of the ‘information’ economy on which the western post-industrial countries rely. Finally in this section, the mass collection of medical data begs a number of question given the privatization of much of the UK’s NHS, including who is able to get access to this data and how will it be used?

Statistical models of the capitalist economy

Before moving to a conclusion, let us consider the widespread use of econometric modelling. Relying on mean values and the rest of the statistics toolkit, quantitative data of doubtful veracity is collected on such variables as inflation, national income and unemployment, early econometricians used inferential techniques to build models typically consisting of regression equations. However, these models were, and remain, unspectacular in terms of both explanatory and predictive power. Nevertheless, with the availability of personal computers and easy to use statistical packages, today economics departments in universities around the world, offer core courses in econometrics. Despite their poor record, the use of these models of markets remains *de rigueur* in government departments, orthodox economics journal articles, financial market dealing rooms, corporate-financed think-tanks and elsewhere. What these models lack in terms of explaining capitalist social relations is made up for by an expanding thicket of algebra and a range of statistical Shibboleths chanted by a caste of econometric high priests.

Rather than being guided by an overarching or totalising approach to capitalism, one that does justice to its interconnectedness, dynamism and internal contradictions, rather like programmed robots, econometricians feed low quality data sets into their SPSS, R and other software in order to develop and test a range of hypotheses concerning economic variables. With this in mind the economist Ronald Coase famously argued: “If you torture the data long enough, it will confess”; quoted in Smith (2016, 5). The hypothesis which stands up to various rule of thumb significance tests applied to the data will be the one that gets published: they “will test many theories but only report the results that are statistically significant. Even if only worthless theories are considered, one out of every twenty tests of worthless theories will be statistically significant. With mountains of data, powerful comput-

ers, and incredible pressure to produce publishable results, untold numbers of worthless theories get tested”; Smith (20). In fact, the few econometrics models that have attracted attention and informed policy have been shown to have involved data manipulation, including unwarranted deleting of outliers and averaging of data, or simply omitting data which did not support their hypothesis.

For example, the doyen of neoliberalism Milton Friedman was caught red-handed manipulating data in an attempt to prove his anti-government, pro-market, argument that inflation was caused by lax monetary policy; see Hendry and Ericsson (1983). Similarly, seeking to bolster their argument that government debt hinders economic growth, a view which continues to inspire the austerity packages of the UK and other administrations, Reinhart and Rogoff were also caught fiddling their data; see Smith (2016). Steven Levitt, bestselling author and Friedman acolyte, when caught data fixing in two academic papers, said “This is personally quite embarrassing because I pride myself on being careful with data”; Smith (69). Given its track record of mediocrity and manipulation, it will come as little surprise to report that, notwithstanding the use of high powered statistics and super-computer technology, not one econometric model was able to predict the coming of the 2007/8 financial meltdown. As Sedlacek (2013, 316-7) points out: only when a clock breaks down “do we find out whether we *really understand how the clock works*...Economists only know how to comment on the economy and fine-tune it, as long as everything is functioning generally well...we completely forgot how clueless economics is in times of crisis”. He sums up the paucity of econometrics: in constructing models we must abstract from the totality of the real world and make simplifying assumptions, but adds, when “applying these models to reality, we must instead look away from the models. We must, so to speak, tear down the scaffolding to see if there is anything left standing beneath it”; (303). Following this look at the pseudo-science that is econometrics, let us move to some concluding remarks.

Concluding remarks: the tyranny of numbers

We have charted the checkered history of career statisticians and noted that some are the hired hands, or software clickers, of the capitalist class. In response to the dubious real world associations of their craft, including the issues surrounding big data and big brother, most stat-

isticians repeat the mantra that their techniques are neutral. As we have noted, ‘neutrality’ seems to mean either ordained by God or inherent in nature. We can take the currently dominant base 10 system of numbers: 0 1 2 3 4 5... and compare it with, to choose one of many, the Roman system: I II III IV V.... Which of these systems are so ordained? The answer is neither of them, or indeed any number system known to humanity. Our current number system is an artifact which only really developed with the onset of banking, accounting and the industrial revolution. Clearly, the Roman system developed to suit the needs of the ruling class of around two thousand years ago and statistics, as we know it, was not part of their agenda; indeed, having no zero, statistics would have been difficult, if not impossible, using their number system. The computation of such artifacts as the mean value, as a near reflex action, has become a reified symbol of oppression in many areas of contemporary life, including educational assessment and the workplace.

We have also noted that in response to this, some “left wing” statisticians have argued that statistics can be subverted and used against the capitalist class. Whilst there is a one-sided truth in this assertion, it is to view statistics from the perspective of critical thinking, to quote a term typically used in academic assessment criteria, and that this is to view statistics from inside its own paradigm. This is not far from a Huff style argument that we can, for example, make political opinion polls more accurate by eliminating bias. This is to miss the point; the functions of such polls are to create working class political passivity by means of the spectacle and legitimise the representative democracy system that is fundamental to global capitalism. More appropriate is to view statistics from the perspective of the 19th century German term *Critique*, which means criticism with a view to creating fundamental social transformation. This involves challenging the tyranny of quantification and creating a world in which human *qualities* are paramount and quantification is confined a place determined by a socioeconomic process in which workplace democracy is central. Instead of the quality *intelligence* being measured by how fast someone can finish a Sudoku puzzle, which is more or less how IQ tests work, we can encourage human creativity in all its manifestations and use it to look after, rather than exploit, each other and act as custodians of our natural environment. Finally, we can note that Facebook made around

\$40 billion profit in 2017; we do not have to live in a world in which the world's largest and most profitable companies are not building houses, making clothes or growing food, but are simply selling data.

Bibliography

- Bernal, M. (1987) *Black Athena: The Afroasiatic Roots of Classical Civilization*, Rutgers University Press.
- Black, E. (2004) *Then They Came for the Gypsies: The Legacy of Death's Calculator*, *The Forward*; July 11th.
- CIA (2011), *Soviet Statisticians' Non-Marxist Views Censured at Methodology Conference of Central Statistical Administration USSR*, <https://www.cia.gov/library/readingroom/document/cia-rdp80-00809a000600360264-9>; accessed 16th April 2018.
- Clegg, B. (2017) *Big Data: How the Information Revolution is Transforming our Lives*, London: Icon.
- Debord, G. (1977, originally published in French in 1967) *Society of the Spectacle*, Exeter: Red and Black.
- Dorling, D. and Simpson, S. editors (1999) *Statistics in Society: The Arithmetic of Politics*, London: Arnold.
- Harding, L. (2014) *The Snowden Files: The Inside Story of the World's Most Wanted Man*, London: Faber & Faber.
- Hegel, G.W.F. (1977) *The Phenomenology of Spirit*, Oxford: OUP.
- Hendry, D. F. and Ericsson, N. R. (1983) "Assertion without empirical basis: an econometric appraisal of monetary trends in the United Kingdom, by Milton Friedman and Anna J. Schwartz", Board of Governors of the Federal Reserve System International Finance Discussion Paper No. 270.
- Hersh, R. (1997) *What is Mathematics, Really?* London: Jonathan Cape.
- Huff, D. (1976) *How to Lie with Statistics*, Harmondsworth: Penguin.
- Irvine, J. et al, editors (1981) *Demystifying Social Statistics*, London; Pluto Press.
- Levy, J. (2013) *A Curious History of Mathematics*, London: André Deutsch.
- Mackenzie, D. (1981) *Eugenics and the Rise of Mathematical Statistics in Britain*, in Irvine et al.
- Marx, K. (1976) *Capital volume 1*, London: Penguin.

- Mayer-Schönerberger, V. and Cukier, K. (2017) *Big Data: The Essential Guide to Work, Life and Learning in the Age of Insight*, London: John Murray.
- Muller, J.Z. (2018) *The Tyranny of Metrics*, Princeton: Princeton University Press.
- Rogerson, B. (2013) *Rogerson's Book of Numbers*, New York: Picador.
- Rose, T. (2015) *The End of Average: How to Succeed in a World that Values Sameness*, London: Penguin.
- Schneier, B. (2016), *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*, New York: Norton.
- Sedlacek, T. (2013) *Economics of Good and Evil*, New York: Oxford University Press.
- Smith, G (2016), *Standard Deviations*, London: Duckworth.
- Suhara, M. (2017) *Russian Industrial Statistics*, available at http://hermes-ir.lib.hitu.ac.jp/rs/bitstream/10086/28396/1/RRC_WP_No66.pdf downloaded on 25th April 2018.
- Ward, J. (2018) *Microsoft, the ID2020 Alliance, universal digital identification and you*. (Online) available at: <https://www.windowscentral.com/microsoft-universal-digital-identification-and-you>; accessed 25th February 2018.
- Young, R.M. (1981) *Why are Figures so Significant? The Role and the Critique of Quantification*. In Irvine, J. et al.