CURRENT SCIENCE

Volume 101 Number 1 10 July 2011

EDITORIAL

Defaming the Dead: History and Statistics

Deconstructing the work of famous scientists, many years after their passing, is a task that has been undertaken by many historians of science and writers of great distinction. A recent editorial in *Nature* (2011, **474**, 419) notes that 'it is impossible to libel the dead, but equally impossible for them to defend themselves'. There is a certain excitement in following the trail of research published decades or even a century ago and discussing evidence for misinterpretation. The thrill heightens when the conclusions drawn from experiments and observations appear to be the result of conscious bias or deliberate tuning of data, to fit a compelling hypothesis. From Newton to Mendel, science's icons have been dissected long after their deaths, by modern historians of science, aided by the benefits of hindsight. In an article that appeared almost forty years ago, entitled 'Newton and the Fudge Factor', Richard Westfall passed judgement: 'Not the least part of the *Principia*'s persuasiveness was its deliberate pretense to a degree of precision quite beyond its legitimate claim. If the *Principia* established the quantitative pattern of modern science, it equally suggested a less sublime truth – that no one can manipulate the fudge factor quite so effectively as the master mathematician himself' (*Science*, 1973, **179**, 751). Fudge factors have been commonplace in the centuries since Newton, as theorists strive to reconcile their results with experimental outcomes. Adjustable parameters are the staple of empirical theoretical approaches. Newton could 'mend the numbers', his physical insights reaffirmed by calculations whose precision, in Westfall's analysis, were often 'brazenly manipulated'.

In 1936, seventy years after Gregor Mendel published his immortal paper which launched the field of genetics, R. A. Fisher, a celebrated statistician, muddied the waters. Mendel's paper, unlike much of the modern literature, provided enough raw data to permit independent analysis. In devoting a great deal of effort to re-analysing Mendel's data, Fisher came to a startling conclusion: '… the data of most, if not all, of the experiments have been falsified so as to agree with Mendel's expectation' (Fisher, R. A., *Ann. Sci*., 1936, **1**, 115; http://digital.library.adelaide.edu/). Fisher's analysis, in turn, was dissected seventy years later leading to the hope that 'against all experience that Fisher's allegation of deliberate falsification can finally be put to rest, because on closer analysis it has proved to be unsupported by convincing evidence' (Hartl, D. L. and Fairbanks, D. J., *Genetics*, 2007, **175**, 975). Scientists are often accused of 'selecting data' to fit a pet hypothesis; errant data points that deviate wildly from expectation are discarded as erroneous. This is a practice that is not uncommon; one that statisticians seem to frown upon. Fisher, however, did not appear to be critical of Mendel's 'data trimming'; rather he seemed to be raising an alarm that some of Mendel's data seemed to fit an expectation much too well. A common saying in experimental laboratories is that sometimes 'data appears to be too good to be true'. Happy accidents do happen, although statisticians may add that the probabilities are low. C. R. Rao, the eminent statistician revisited the Fisher analysis noting that the overall probability of agreement between Mendel's data and his expectation 'is 7/100000, which is very small'. He goes on to quote Fisher: 'Although no explanation can be expected to be satisfactory it remains the possibility among others that Mendel was deceived by some assistant who knew too well what was expected' (Rao, C. R., *Statistics and Truth*, World Scientific, Singapore, 1997, pp. 78–79). Hartl and Fairbanks suggest that Fisher may have been misled by his reading of the traits that Mendel scored in his experiments. There are other explanations too. While Mendel's work remains one of the conceptual pillars on which modern biology rests, allegations once made tend to resurface. In their title, Hartl and Fairbanks emphasize the fact that 'mud sticks'. They pay the highest tributes to the manner in which Mendel presented his raw data, noting that his 'celebrated paper … is an inexhaustible source of inspiration and controversy for each succeeding generation of geneticists and historians of genetics'. In their assessment, 'Mendel's paper appears to reflect the author's simplicity, modesty, and guilelessness'. C. R. Rao credits Mendel with 'introducing the *indeterministic paradigm* in the history of science … by observing data subject to chance fluctuations'; a precursor to modern evolutionary theory which has been described as 'a mixture of chance and necessity – chance at the level of variation and necessity in the working of selection' (pp. 35–36). Newton and Mendel are iconic examples of creativity; scientists whose instincts and insights seem to enable them to slice away irrelevant and confusing data. To the purist, steeped in the modern methods of data analysis, there is a disturbing hint of picking and choosing data; fudging a calculation by mathematical sleight of

hand. Reading about the attempts to deconstruct Newton and Mendel, I was reminded of how chance often showers its blessings on those who are singularly talented. Football fans will recall Diego Maradona's famous goal for Argentina in a World Cup quarter final against England in 1986, when the 'hand of God' propelled the Argentinians to a famous victory, in a clash that took place not long after the Falklands War between the two countries.

Few writers have matched Stephen Jay Gould in his abilities to revisit famous successes, failures and controversies in science. In 1987, Gould resurrected an old controversy in paleontology, a field replete with many famous disputes. In turning to the fossil of the Archaeopteryx, 'an evolutionary intermediate, a transitional form between reptiles and birds', Gould focused on one of Victorian England's famous fueds; the running battle between Richard Owen and Thomas Huxley (*New Scientist*, 12 March 1987, pp. 32–36). The fossil was a prized acquisition for the British Museum, then headed by Richard Owen. Darwin's theories had just been published and here was a fossil likely to link two major biological lineages. Unfortunately, Owen interpreted the orientation of the fossil on the slab incorrectly, an error that left him open to Huxley's merciless criticism. Huxley had scored another victory over his foe. But Gould points out that 'the fullness of history often drowns an immediate triumph in irony. Although he muffed his description, we now realise that Owen was right and Huxley wrong, in the central aspect of their general interpretations'. Gould's 1987 essay entitled 'The Fossil Fraud that Never Was', was sparked by a claim by Fred Hoyle that the Archaeopteryx fossil is a forgery, 'manufactured from modern feathers pressed into a layer of artificial cement painted atop the genuine skeleton of a small dinosaur'. Gould describes Hoyle as an 'eminent cosmologist of years past and more recent supporter of uncommon causes' and proceeds to demolish Hoyle's attempt to paint Owen as the perpetrator of a major fraud. Gould has advice for historians who revel in deconstruction: 'I do not advocate an unthinking ancestor worship, but scholars do make a tacit pledge when they enter this exacting profession – to honour the struggles of those who have gone before and to treat arguments with respect and integrity'. The Hoyle–Wickramasinghe attack on Richard Owen would have been considered libellous if it had not been made nearly a century after Owen's death. A review of the book *Archaeopteryx, The Primordial Bird* by Fred Hoyle and Chandra Wickramasinghe (Christopher Davis, 1987) describes it as containing 'demonstrable falsehoods, as well as hardly imaginable calumnies of persons unable to defend themselves' (Halstead, B., *New Scientist*, 10 September 1987, p. 70). Indeed many attempts at deconstructing past scientific discoveries often degenerate into an obsessive attempt to defame long dead scientists.

The most recent example is an analysis which suggests that Stephen Gould's 1978 dissection of Samuel Morton's work on cranial capacity was flawed (Lewis, J. E. *et al.*, *PLoS Biology*, 2011, **9**, e1001071). Morton, a physician, published a measurement of cranial capacity, based on a collection of nearly 1000 human skulls in 1839 and 1849. Cranial capacity was thought to be a rough indicator of overall intelligence in Morton's study. Human racial differences appeared to be manifested in his measurements, leading inescapably to a ranking of races. Morton's work passed into obscurity until Gould resurrected it 127 years after Morton's death. In a paper in 1978, intriguingly subtitled 'Unconscious manipulation of data may be a scientific norm', Gould sets out to demonstrate that Morton's biases 'produced the results anticipated in an age when few Caucasians doubted their innate superiority' (*Science*, 1978, **200**, 503). Gould quotes extensively from Morton's writings which have a tone that is unabashedly racist, but must be viewed in the specific cultural and historical context of the first half of the 19th century in America. Curiously, Morton's measurements included as many as '17 Hindu skulls' leading him to note that 'the skulls of these people are probably smaller than those of any other existing nation'. Gould argues that Morton chose his data selectively, made convenient omissions and points to 'slips' and 'miscalculations'. He concludes, when the 'data are properly reinterpreted, all races have approximately equal capacities'. His judgement of Morton is charitable: 'I find no indication of fraud or conscious manipulation'. In reflecting on Morton's work, Gould notes 'that lying with statistics is easier than fudging an experiment'. He adds: 'Unconscious or dimly perceived finagling is probably endemic in science, since scientists are human beings rooted in cultural contexts, not automatons directed toward external truth'. Three years after his analysis in *Science*, Gould used Morton's work to title his book, *The Mismeasure of Man* (W. W. Norton, New York, 1981). Three decades later, Lewis *et al.* 'falsify Gould's hypothesis that Morton manipulated his data to conform with his *a priori* views'. Readers of this paper in *PLoS Biology* may find parallels to the Hartl–Fairbanks reanalysis of the Mendel–Fisher controversy. Ironically, Lewis *et al.* conclude that 'Gould's own analysis of Morton is likely the stronger example of bias influencing results'.

Biases and prejudices are an integral part of human nature. Experiments must support hypotheses, theory must edge towards agreement with experiment; these are the pressures of science. Gould was right when he argued that 'finagling is not a disease'. He was even more right when he added, 'the only palliations … are vigilance and scrutiny'. To buttress my limited knowledge of statistics and statisticians, I had turned to C. R. Rao's *Statistics and Truth*. Here I found a quote that seemed apt as a conclusion:

'*All knowledge is, in final analysis, history*
All sciences are, in the abstract, mathematics
All judgements are, in their rationale, statistics.'

P. Balaram