Learning the Lessons of History?

Scientific Fraud in the Low Countries

[VITTORIO BUSATO]

'If the researcher first strives to the best of their ability for objectivity and pure reasoning, and secondly adheres to the prescripts of honest research and open publication as described in this book, then there is no room for confusion. Others may then repeat his work if they wish and/or may identify where, despite his best efforts, influences from ideologies other than that of objective science have crept into the researcher's definitions, or where he has committed methodological errors.'

This quotation is more than fifty years old and comes from the book *Methodologie. Grondslagen van onderzoek en denken in de gedragswetenschappen* (Den Haag, 1961) by Adriaan de Groot (1914-2006). The book was translated in 1969 as *Methodology. Foundations of inference and research in the behavioral sciences*, but all quotations in this essay are translated from the Dutch original. De Groot spent a long time as a professor at the University of Amsterdam, and a shorter period at the University of Groningen. Many still see him as the most influential Dutch psychologist of all time. His 1946 doctoral thesis, *Thought and choice in chess* (1965), is regarded as the intellectual precursor of cognitive psychology. Another significant, more national contribution was his encouragement of social scientists in general and psychologists in particular to provide empirical support for their theories. In *Methodologie* – more than fifty years ago, remember (the Dutch-language edition dates from 1961) – he set out in crystal clear fashion how that could be done as fairly as possible.

Generations of social science students have been educated with his empirical cycle and related research maxim: 'If I know something, I can predict something; if I cannot predict anything, then I know nothing.' This empirical cycle begins with *observation*: the systematic collecting of empirical facts and the formulating of hypotheses. The second phase is *induction*, in which the hypotheses are formulated more precisely. Next comes the phase of *deduction*: the formulation of specific predictions based on those hypotheses. These predictions are then empirically *tested* using new empirical material. In the fifth phase, the results are *evaluated* for their theoretical validity, after which the cycle starts again from the beginning.

De Groot argues that it is crucial during this process to draw and maintain an explicit distinction between exploratory and verification research. Hypotheses, he argued, must be formulated *in advance* and tested using *new* data. 'Anyone who imagines exploration in reporting as verification research by acting as if the hypothesis had already been formulated precisely before the research began – something which is sadly all too easy to do – is guilty of a serious infringement against the social ethics of science. In the 'open' communication between scientists, it is expected that such misrepresentations will not occur.'

The purpose of exploratory research, De Groot stresses, is to build hypotheses. Hypotheses are not tested during exploratory research; as with verification research, they are not precisely formulated in advance, but are simply explored in order to formulate more precise hypotheses in line with pre-existing theoretical findings. 'The notion of 'exploration' too often turns out to be a euphemism for unnecessary contamination in a piece of research which would have been far better if it had had a systematic, objectively descriptive design. (...) The consequence is that what was initially intended as verification research, but in reality is a poorly executed project, is presented as 'exploratory research' as a last remedy for methodological shortcoming.'

According to De Groot, exploratory research is *preliminary research*. If it is not followed up by precise theory and/or hypothesis formulation and testing, then it is virtually useless. 'Mixed', theoretically based investigations must be subjected to the requirement cited earlier, namely that the researcher *maintains a clear separation between the different forms and procedures*. The meaning of the level of significance, for example, depends greatly on whether we are dealing with verification or exploratory research. For example, if we continue to explore until we find something 'significant', the preselection means this is no longer significant in a statistical sense.'

The fraudulent inventor

Do we really learn the lessons of history? Judging from a number of recent cases of scientific fraud in the Low Countries, there is some cause to doubt this. If 'researchers' such as Diederik Stapel (Tilburg University), Dirk Smeesters (Erasmus University Rotterdam), Don Poldermans (also Erasmus), Mart Bax (VU University Amsterdam), Peter Paul Rijpkema (University of Amsterdam), Patrick van Calster (University of Groningen) and a Flemish professor of rheumatology at Leiden University have studied *Methodologie* at all, it is clear that they have paid scant regard to that empirical evidence base and to prescriptions relating to honest research and open publication – to say nothing of the explicit distinction between exploration and verification research.

What were their transgressions, again? The social psychologist Stapel turned out to have invented much of his oeuvre; he climbed rapidly into the international top ten of the greatest scientific fraudsters. His Flemish colleague and professor of consumer behaviour Dirk Smeesters, also a social psychologist, was caught out manipulating data in three articles that have since been withdrawn.

The internist and university professor Don Poldermans fabricated data for around two hundred patients. Mart Bax, a professor of political anthropology, was charged with scientific misconduct long after retirement (though without consequence, as if his fraudulent inventions had passed their statute of limitations); among other things he was accused of citing non-existent publications, falsified achievements and distinctions that had never been awarded in official documents, and basing his publications on unverifiable source references.

Law professor Peter Paul Rijpkema was accused of plagiarising large sections of a book published in his name from a book written by his predecessor, without crediting him; unlike Stapel, Smeesters, Poldermans and Bax, Rijpkema's actions were ultimately not classed as infringements of scientific integrity, but as extreme lack of care. The transcriptions by Professor of Criminology Van Calster, by contrast, did cost him his job; in 2005 the Vrije Universiteit Brussel stripped him of his doctorate because large tracts of his thesis had been taken without citation of sources from a widely used management manual; this proved to be the hair that broke the camel's back for his Groningen employer.

And what of Annemie Schuerwegh, the rheumatologist? She was dismissed after colleagues discovered fraud in her laboratory research. It turned out that A.S. had been creeping into the laboratory at night in order to manipulate patients' blood samples so as to disguise the fact that a test she herself had designed was not fit for purpose. Fortunately her patients, like those under Poldermans, were never in danger.

Pressure to publish

What motivates scientists, who know that they are at great risk of being unmasked through the self-cleansing power (at least that is its intent) of peer review, nonetheless to commit fraud? Is it the constant pressure on academics to publish, as Stapel himself stated in a bid to explain his fraud? Should the cause of this fraud perhaps be sought in too much vanity, egotism or narcissism, prompting those concerned to use fraudulent means if necessary to secure a presence in journals with the highest *impact factor*, those magic words in the world of science today?[1]

This pressure to publish, the rat race it engenders and the emphasis on quantity is sometimes lamented, for example in the Flemish newspaper *De Morgen*, by a large group of scientists as a response to the fact that the Flemish government distributes funding to academic institutions primarily on the basis of the numbers of publications and numbers of students and doctorates (the notorious output funding). This, it was argued, increases the temptation to put empirical reality in a slightly more favourable light in order to increase the chance of publication – all the more so because journal editors are much more inclined to publish 'significant' results.

This pressure, or perhaps more accurately the market-based thinking of academic institutions, can also give rise to perverse incentives. If universities are funded partly on the basis of the number of graduates, the danger of academic inflation is not far away; it is easy to adjust the standard of an examination. If universities receive more money for producing more publications, we should not be surprised that management will focus on raising the number of publications. One professor, who had been invited to apply for a professorship – at Tilburg, ironically – told me for example of being tempted 'by bonuses of € 5,000', on top of the salary, for every publication that appeared in a top journal – though it must be said that the party in question regarded this practice as unique to that university.

It is kitchen sink psychology, of course, but it seems equally likely that narcissism cannot be ruled out as a potential (partial) cause of scientific fraud.

The scientific philosopher Sir Karl Popper (1902-1994) – a close acquaintance of De Groot, as it happens – once said that science is not so much about *who* says something as about *what* is said; that appears to be less and less the case in today's scientific judgement-by-results culture. Take a quick look on the Internet at CVs of academic psychologists; many of them report the impact figures of the journals in which they publish, as well as the number of times their articles have been cited – figures which naturally change over time. Is this kind of self-congratulation the way to express the fact that one 'counts'? Is it the best way of expressing the real impact in relation to the task of psychology and psychologists, namely to understand and help others?

Wherever the truth may lie, pressure to publish or excessive vanity can never be a single explanation for scientific fraud. There are after all very many scientists, vain or otherwise, who struggle under the perceived pressure to publish but who are not tempted into fraud. The reality is that people in general, and therefore also scientists, are prone to cheating, to engage in dodgy business, to make mistakes, to lie, and in some cases to cross the line.

Grey area

Different categories of infringements of integrity are recognised in the world of science. Inventing data stands firmly in top place, followed by plagiarism and extensive manipulation of data. Below this, however, is a much greyer area.

Practices such as omitting data that do not fit the researcher's purpose, adjusting statistical analyses so that the results turn out more favourably, recycling or splitting research results purely in order to reach more publications, adapting and 'sexing up' hypotheses retrospectively in order to obtain results that are significant – did I already mention that De Groot highlighted such practices more than fifty years ago? – are all today regarded as questionable research practices.[2] It is by no means always clear how deliberately these practices have been applied. For example, if a researcher freely admits that they have eliminated an outlier from the data set (say, for example, a subject who deliberately sabotages a research project) then, unlike a researcher who does this 'clandestinely', they are not guilty of any wrongdoing at all.

How common are such questionable research practices? According to a recent study by Harvard Business School, one in ten psychologists may at some time have improperly forged data.[3] According to a more recent doctoral thesis, half the scientific publications in the field of experimental psychology contain statistical anomalies, ranging from rounding data up or down to sanitizing data to obtain a more favourable outcome. Earlier research led by Jelte Wicherts, currently senior lecturer at the Department of Methodology and Statistics at Tilburg University and initiator of the recently founded *Journal of Open Psychology Data*, suggested that psychologists who do not publish their data may also have something to hide. By way of illustration, a few years ago he and colleagues published a report of a remarkable survey in the journal *American Psychologist*.[4] Of the authors contacted who had published in the last two issues in 2004 of the high-impact journals *Journal of Personality and Social Psychology, Developmental Psychology, Journal of Consulting and Clinical Psychology* and *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, only a quarter (!) made their data available for reanalysis. That is interesting, because the ethical guidelines of the *American Psychological Association* (APA) impose a requirement on researchers to make their data available to colleagues for at least five years. Even more remarkably, in another study Wicherts et al. found that researchers who do not make their data available also make noticeably more statistical errors.[5]

It remains unclear how systematically psychologists fiddle with the statistics. Would those psychologists admit their unfair play if they were not allowed to remain anonymous? How many of them apply their unfair practices or even commit fraud in a cleverer and more subtle way and so slip through the net? It is difficult to establish precisely how often fraud occurs in psychology, even though the public at large may think that since the Stapel scandal the incidences of fraud are piling up and strict controls are absolutely essential. But in a profession that is characterised more than many others by mutual trust, it is not possible to exercise total control, as if researchers were potential drug mules on a flight from Curaçao to Amsterdam.

Back to square one

This does not alter the fact that journals of psychology are increasingly filled with calls to restore transparency and fair play in research, among other things by encouraging replication, sharing and publicising data and pre-registering experiments (where scientists state in advance what they plan to research, how they intend to do so and which conclusions they do and do not wish to be able to draw from their research). It is however questionable how new these initiatives are and whether devoting much more attention to the work of an old master in the profession such as De Groot might not be a much simpler remedy.

As stated earlier, journals publish far more significant than non-significant results. Replication studies have virtually no chance of being published. This is strange, because a replicated effect strengthens the effect originally found and a non-replicated effect places that effect in empirical perspective. The Open Science Framework is currently setting up any number of replication initiatives which are intended to form part of the scientific cycle with the aim of delivering more robust knowledge; one study does not after all constitute a body of research. This is indisputably laudable, but more than fifty years ago De Groot was already expressing surprise at how sporadically replication studies appeared: 'And if they are carried out, the results – entirely without justification– are often not published, especially if they are negative.'

De Groot also had something to say long ago about that transparency and sharing of data which Wicherts et al., among others, so laudably and loudly proclaim; read the opening quotation in this article once again. And fair is fair, researchers who call for pre-registration will also (yet again) need to turn to De Groot: 'The most detailed possible advance description of the verification (or experimental) design is in any event strongly advisable.' (De Groot, *Methodologie*)

Fair play in research also has to do with giving credits to those who historically deserve them; that is without doubt the best way to genuinely learn from the lessons of history. It is therefore time in the world of science, and definitely in the world of psychology, to pause for a moment. The kind of artisanal professionalism that De Groot described more than fifty years ago in Methodologie should once again become compulsory reading in scientific training. Compulsory *Methodologie* refresher courses should also be organised for doctoral students, postdoctoral researchers and professors (and unquestionably also for all social psychologists).

Once this is all solidly in place, there will be a significant reduction in the current excessive variation in knowledge about methodology and statistics among psychologists. I would then venture to predict that not only will questionable research practices become a thing of the past, but that it will also become clear that psychologists still know very little indeed. Much of the knowledge in the world of psychology, partly because of this unfair play, is based on theoretical quicksand. Making the study of *Methodologie* compulsory would help ensure that solid foundations are laid before erecting new structures.

And I would also dare to predict that a side-effect will be that the emphasis on quantity in science (at least in psychology) will decline automatically. Because if I know anything thanks to De Groot's masterful book, then it is that measurement may provide knowledge, but simply counting adds up to nothing.

Postscript

A brief final word about impact. A friend of mine is a widow; her son and daughter lost their father at the ages of six and eight, respectively. Her daughter is now fourteen. It recently struck me that she was so changed, as if her shyness and uncertainty had suddenly given way to an almost adult self-awareness. What had happened? It transpired that my friend's daughter had been to a bereavement counselling weekend for children who had lost a parent or sibling. That weekend was led by a psychologist from whom she said that she had learned an enormous amount about coming to terms with her emotions concerning the loss of her father and sharing that with others. *That* is impact. Would that counsellor also boast about it in his or her CV? ■

Translated by Julian Ross

NOTES

1) An impact factor is calculated based on the average number of citations of all articles published in a journal within a period of two years. The higher the impact factor, the higher the scientific prestige of a journal. Better journals are read by more scientists. That increases the chance that articles from that journal will be cited. That in turn increases the impact of the journal – and so on, and so on.
2) L.K. JOHN, G.F. LOEWENSTEIN & D. PRELEC, 'Measuring the Prevalence of Questionable Research Practices with Incentives for Truth-Telling', in: *Psychological Science*, 23, 2012, pp. 524-532. See also: M. Bakker, *Good science, bad science. Questioning research practices in psychological research, academic thesis*, University of Amsterdam, 2014.
3) See first reference in note 2.
4) J.M. WICHERTS, D. BORSBOOM, J. KATS & D. MOLENAAR, 'The poor availability of psychological research data for reanalysis', in: *American Psychologist*, 61, 2006, pp. 726-728.
5) J.M. WICHERTS, M. BAKKER & D. MOLENAAR, 'Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results', *PLoS ONE*, 2011.doi: 10.1371/journal.pone.0026828.