Accumulating Evidence of Malfunctioning Contractual Government Machinery

Mervyn Stone

Evidence is presented from seven scenarios in which the author has been engaged in recent years, that the machinery of government has two related weaknesses—in its neglect of analytical statistical thinking and in the contractual procedure for using external expertise. Their remedy is outlined in illuminating historical context.

Mervyn Stone is Emeritus Professor of Statistics at University College London

The Spottiswoode Report “Improving Police Performance” for the Treasury’s Public Services Productivity Panel (Spottiswoode. 2000) is a good example of a “systematic application of thought” (Haldane. 1918. para.12) about an issue of public importance that came to nothing in a technical cul-de-sac. A postmortem in this journal (Stone. 2002b) ended with an explanation of how that outcome could have been foreseen, and a suggestion as to how similar outcomes might be avoided in the future—by changes in the contractual machinery that government departments use when they employ external expertise.

Here, I present evidence that, in the main, these changes have been neither made nor envisaged. The evidence comes from a number of adventitious and necessarily personal engagements that cannot be depersonalised without weakening their cumulative message. The next section will therefore not refrain from uncovering superficially sensitive matters that illuminate the way traditionally secretive government departments and agencies handle scientific and technical expertise from outside. The final section returns to the general lessons that still need to be learnt.

Seven personal engagements

1: Well before the Spottiswoode Report, I had come across evidence that things were not what they should be. In the last years of the Thatcher government, I was contracted to review the statistical work of the water authorities and companies in the “K-factor” case they were making for the charges needed after privatisation to replace all the clogging and leaking water pipes. Many millions of pounds depended on how many would have to be replaced—and this was very costly to estimate. The quality of the work varied greatly. Some of it was done by legal firms that claimed to have the necessary expertise. Some of it was “classical” statistical method (unbiased estimates etc) based on measurements on pipes in the ground, but some of it was Bayesian, allowing an interested party to adjust the measurements with prior knowledge based on in-house expertise. Even if it had been taken seriously, my single overall report—to the civil engineering company that had been contracted by government to handle things—could hardly have been adequate as a peer review of all the reports that were influential in sanctifying the financial arrangements of the privatisation.

2: Such concern can be legitimately extended to the work of agencies of government, among which the European Commission (EC) should now be included, since its decisions affect us all and are partly under British government
control. In 1998, I wanted to know more about the chances of Europe-wide implementation of some EC-financed research that was making impressive claims about the reduction in road accidents to be expected if we adopted “daytime running lights” for vehicles. Eventually, I had to contact Edith Cresson’s Directorate-General for Science, Research & Development to learn more—before publishing a refutation (Stone, 1999) of probabilistic arguments suggesting Cresson’s dentist may have had a hand in the research.

3: Suspicion, that both here and abroad the organisation of such scientific work was not as it should be, hardened to certainty with the events surrounding the Spottiswoode Report. My unpremeditated involvement in these events was in response to an appeal to some mathematicians and statisticians for independent advice about the intended application of Data Envelopment Analysis (DEA). The appeal came from a civil servant working on short-term contract for the Treasury/Home Office Panel, whose intuition questioned the claims being made for DEA. The Spottiswoode Report’s attempt to exclude value judgements about individual police force outputs from its recommendations for measuring force efficiency should challenge anyone’s intuition about the social relevance of DEA.

From the very start—when DEA has to nominate the wide range of outputs from which, taking police force costs into account, it will then derive the alleged “efficiencies”—one has to ask how it can make that selection without value judgements of some sort. But the greater challenge to good sense as well as intuition comes when we see the kind of thing that DEA can do with data it might get from individual police forces. The fuller story is in Stone 2002b. Here, just one example will suffice. If any one of the 43 police forces were to concentrate its resources on just one of the DEA-nominated outputs and hence become the force with the best performance in that output, it would be free to put up a brass plate with an inscribed “efficiency” score of 100%. Truly “By their fruits ye shall know them [techniques such as DEA]!”! Coleridge’s Christabel says it all, more precisely, with “Carved with figures strange & sweet/All carved out of the carver’s brain”

Features of the case not previously documented are:

(a) reliance by the HMT/HO panel on academics already over-committed to DEA;
(b) dismissal of the concerned civil servant instead of taking on board the independent views she brought to the attention of the panel;
(c) costly use of PriceWaterhouseCoopers (PWC) to sanctify a much-delayed face-saving relegation of the DEA approach;
(d) Home Office’s refusal to let the Royal Statistical Society audit the PWC report at the request of the Association of Police Authorities.

4: One year after the publication (Stone, 2002a) that I hoped would drive a stake through the heart of DEA, I was asked by the Department for Education & Science (DfES) to tender for a contract to study how to do for universities what Spottiswoode had tried to do for police forces. There in the middle of the Invitation to Tender document
was the apparently resuscitated acronym! DEA was still being considered by some civil servants in DfES and other government departments despite its relegation by HMT/HO. However, there must have been some doubts in DfES about it, otherwise I would not have been asked to tender. I had to decline the invitation to tender and the contract was given to academics keen to put their DEA expertise to work. However, with another critic of DEA I was invited to comment freely on their report. At the time of writing, I am looking at the DfES website to see what has finally emerged.

5: It was lay intuition about an official publication from the hinterland of Home Office statistics that led to a critical research report (Stone, 2003) on net immigration from the new EU countries. Ex-ambassador to Saudi Arabia, Sir Andrew Green, was grinding his axe and wanted statistical blessing before issuing yet another Migration Watch contribution to the public debate about immigration. Readers are free to download my report to see whether they can agree with the stoutly asserted and oft-repeated Home Office claim, misrepresenting the itself dubious work of a single academic research group commissioned by the Immigration and Nationality Directorate—that there was no reason to question their forecast of a net immigration of between 5,000 and 13,000 when the EU went from 15 to 25. This case incidentally revealed the difficulty of keeping such matters academic. My report had also been posted on the Civitas website and, when Civitas did a Press Release about the report, the Sunday Times and Daily Mail responded. I had one minute on Channel Five and two minutes on News 24 to tell the world all about it.

6: An unpremeditated engagement with the Department of Health (DoH) stems from another layperson’s statistical intuition. Joan Davis, chairman of North West London’s NHS-monitoring organization Community Voice, had previously persuaded three Royal Statistical Society fellows to support her objections to some poor statistical work in aid of closure of Mount Vernon Cancer Centre by the NHS Eastern Region Cancer Registry. Her second appeal for help was in response to the public health report of her Primary Care Trust (PCT) that mixed great concern for health inequalities and half-baked statistical measures aimed at correcting them (Pickles, 2003). Thinking I could do little at the intensely political local level, I was drawn into the equally political issue of how DoH divides a pot of billions among the 304 PCTs of England. Currently this allocation is forcefully guided by a formula devised in Scotland and derived from a multitude of proxy variables—rather than by direct measurement of health need. The formula bears little relationship to the one that it replaces that had been devised by a different team of health economists at the University of York—and even less relationship to those that have been used, one after the other, since the 1970s. The whole matter is covered in a research report (Stone & Galbraith, 2004).

7: The world-famous Transport Research Laboratory was an executive agency of government until 1996 when it was privatised. From then it became one of many competitors for contracts with the Department of Transport. The management of the Safety Camera Partnership program—the installation of speed cameras at accident black spots in selected police force areas—was entrusted to PA Consulting Ltd (the private sector firm involved in the management
of the work leading up to the Spottiswoode Report). The three-year report on the changes in vehicle speeds and accident incidence in the vicinity of the cameras was released from an election embargo the day before a Radio 4 Today Programme tribunal on the arguments for and against speed cameras and road humps. The Secretary of State for Transport claimed that the cameras had reduced accidents by 40% and saved over 100 lives. I acted as adjudicator for the tribunal and my report (Stone, 2004) was used in broadcasts in the following week. The absence of a proper scientific base for the claims made for the results of the partnership program has now been pointed out by a number of scientists who have published peer-reviewed studies of the influence of speed cameras. The consensus seems to be that claims made without such a base may be destructive of public trust in the machinery of government.

**General lessons in historical context**

Does the greater bulk and complexity of government activity in 2004 mean that we need not pay much attention to a slim report on the machinery of government written towards the end of the First World War (Haldane, 1918)? It might be better to ask whether we can afford not to try to understand why the defining objective of that report has no current prominence on any significant scale. That objective was the “systematic application of thought” that Haldane implicitly assumed would be successful. If my evidence of failures in statistical and “operational research” thinking is representative of current practice in some wider domain—one that would include the as yet inadequately documented government IT fiascos—we must conclude that the Haldane program has bequeathed no useful legacy in that domain. Understanding why it has not done so may give a clue to present remedy.

As Secretary of State for War 1905-12, Haldane had used “thinking”—his own and that of others—and scientific method in the shape of economically-constrained manpower planning, to prepare the British army for the coming war with Germany. One year into that war saw the creation of the Department of Scientific and Industrial Research (DSIR) under a minister without portfolio. In para 12 of Part I of his report, Haldane referred to the need that had thereby been recognised, and added that “there is in our opinion good reason for extending what has been done here to other fields in which thinking is required in aid of administration”.

Haldane’s 1918 Report enunciated a related principle—that in the supervision of non-military research of “general use”, responsibility to Parliament should be in the hands of a minister who is “in normal times free from any serious pressure of administrative duties, and is immune from any suspicion of being biased by administrative considerations against the application of the results of the research.” This principle became the basis of the “arm’s length” or autonomous practices of the research councils. Haldane wanted DSIR to be given a broader remit as a Department of Intelligence & Research (DIR) ranking fourth after Treasury, Defence and Foreign Affairs. With an Olympian detachment that must have either impressed or irritated his colleagues, Haldane saw DIR as a body that could help an administrative department to deal with any problem that may require
“protracted research, involving many kinds of scientific considerations, and may begin to call for the application of the united knowledge of the Department in which [an investigation] originated, and of [DIR]. What may be needed in these cases is a complex, prolonged, and many-sided enquiry, directed not merely to the solution of the particular problem which confronts the administrative Department, but to the attainment of results which may have a profound effect, it may be only in the distant future, both upon pure and upon applied knowledge.”

Haldane did not explain how such liaison between DIR and administrative department could be guaranteed whenever DIR thought it to be necessary. In the event, DIR never materialised. Between 1918 and 1947, DSIR absorbed or created 14 specialised research establishments: for example the Road Research Laboratory from the Ministry of Transport in 1933. Left uncreated was any Statistical Thinking Laboratory whose tasks might have come close to much of what Haldane had in mind. (In para.64 he had listed “the methods of statistical enquiry and the presentation of statistical results” as part of “a generous conception of the scope to be assigned to Intelligence and Research work for general use”.)

The years from 1918 to 1947 almost precisely match the period of formative development elsewhere of the statistical methods that have played, by recent misapplication, a central role in the seven engagements listed above—in stark contrast to the successful role that the same methods have played outside government. The reason for the seven failures (and presumably for others) has two components. The first concerns the relative status within government of statistical and economic science (I here restore the “science” to economics that was removed by Keith Joseph when he changed the Social Science of ESSRC into the Economic and Social of ESRC). The second component is part of the Rothschild legacy. Both parts of the machinery of government are repairable.

Before the Second World War, the profession of statistics had been a departmental matter, dealing with special needs such as censuses and trade figures, whereas the profession of economics had played a more central role in the shape of the Chief Economic Advisor and the Economic Advisory Council. The latter had replaced the shorter-lived Committee of Civil Research that in 1925 had been

“...charged with the duty of giving connected forthought from a central standpoint to the development of economic, scientific and statistical research in relation to civil policy and administration...”

My italics reveal the influence that Haldane was still exercising!

The Second War demanded central coordination and innovation: economic advisers were appointed to many departments and a Central Statistical Office (CSO) was set up in the Cabinet Office. In 1946, statisticians became a “specialist class” whereas the economists stayed within the Executive/Administrative hierarchy. Since then, statis-
ticians in government can almost be said to have been the non-analytical fieldworkers for economic analysis (an echo of the now-abandoned motto on the wheatsheaf logo of the Royal Statistical Society—"aliis exterendum" freely translatable as "to be threshed out by others"). In 1996 the CSO was integrated with the Office of Population & Censuses as the Office for National Statistics to be an executive agency of the Treasury! It was the Treasury that saw the problem of measuring police force efficiency as one for economists to resolve, and approved the trust that PA Consulting put in the hands of an agricultural economist in the preparation of the Spottiswoode Report.

In 1971, Lord Rothschild had the ear of the then Secretary of State for Education and Science and, for applied research, engineered a weakening of the Haldane principle. A substantial amount of research previously supported by the research councils was taken back (with some of the research council money) into departments where it was organised on a “customer-contractor” basis: “The customer says what he wants; the contractor does it (if he can); and the customer pays.” With that, the stage was set for the malfunctioning revealed by our seven engagements. The trend set in the Thatcher/Major era has been vigorously developed in the Blair era. Consider again our engagement 5. It was the Home Office’s Immigration & Advisory Directorate that commissioned the study. Had it come as a detailed proposal to the Economic and Social Research Council for the kind of peer review that I witnessed when I served on its pre-Rothschild predecessor, the Social Science Research Council, it is doubtful that the findings could have been abused as they have been by HO.

If we accept that ministers are unlikely to relinquish the influence they now exercise over the direction of research and the presentation of its results, is there anything that can be done to get more quality into the science? Presentation could be covered by a ministerial code of practice. For the science, the central issue is that, faced with a technical problem beyond departmental competence, some civil servant usually below administrative grade has to formulate the problem, dress it up as an Invitation to Tender, and then somehow ensure that the successful tenderer is competent to do the job. Think about that in the context of any of our engagements 3 to 7! It is a real problem when experts disagree. Mazur (1973) suggested a procedure that would “allow the disagreeing experts to confront each other as adversaries before a panel of judges”. That may have been what happened in the final stages of engagement 3 at the Home Office. My recommendation (Stone, 2002b) is that, in the early stages of any forethought, experts should be kept apart to maximise their independence and that their number should be greatly enlarged and individually motivated to offer a disinterested opinion by, for example, a governmental contribution to their organisation’s tea-club.

Conclusion

Responsibility for remedying our failure to get things right has to be widespread. As Page (2001) notes:

“It would be hopelessly pious, and run counter to what we know about how politics and the media
work to expect politicians and the press to develop an appetite for the sort of detail that makes up the world of everyday politics...there is a limit to the number of conflicts in which the public can remain interested—when you find a new fight to observe, you tend to lose interest in an older one. However, this does not excuse the extreme reluctance of most observers, academic and otherwise, as well as politicians to monitor what is going on in the world.”

Interest in “what is going on” (i.e. reality) comes up against the wisdom in T. S. Eliot’s observation in the first of Four Quartets that “human kind cannot bear very much reality”. The details of some of the above engagements reflect the recent claim of an adviser to the US president that America “now creates its own realities”. The physicist Richard Feynman met the phenomenon when he investigated the failure of the O-rings in the Challenger shuttle disaster and concluded that “reality must take precedence over public relations, for nature cannot be fooled”. Before the First World War, Haldane had over-ruled the unrealities of British army disorganization to prepare for the coming battle. Today, a respect for reality in the management of public expenditure is increasingly urgent at a time when Britain’s main global competitors, China and India, do not suffer the kind of maladministration revealed by our engagement 6 in the disposal of £50B in one year.

References